

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Takanori OKURA
Kakuji TORIGOE
Masahi KURIMOTO
- (ii) TITLE OF INVENTION: GENOMIC DNA ENCODING A POLYPEPTIDE CAPABLE OF
INDUCING THE PRODUCTION OF INTERFERON- γ
- (iii) NUMBER OF SEQUENCES: 35
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: BROWDY AND NEIMARK
(B) STREET: 419 Seventh Street, N.W., Suite 300
(C) CITY: Washington
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20004
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: Patent In Release #1.0, Version #1.30
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: JP 185,305/96
(B) FILING DATE: 27-JUN-1996
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: BROWDY, Roger L.
(B) REGISTRATION NUMBER: 25,618
(C) REFERENCE/DOCKET NUMBER: OKURA=1
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202-628-5197
(B) TELEFAX: 202-737-3528

(2) INFORMATION FOR SEQ ID NO: 1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 157 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

```

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile Arg Asn Leu Asn
1      5      10      15
Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp
20      25      30
Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg Thr Ile Phe Ile
35      40      45
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile
50      55      60
Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys Glu Asn Lys Ile
65      70      75      80
Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys
85      90      95
Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys
100     105     110
Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu
115     120     125

```

Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu
 130 135 140
 Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu Asp
 145 150 155

(2) INFORMATION FOR SEQ ID NO: 2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1120 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: cDNA to mRNA
- (iii) HYPOTHETICAL: No
- (iv) ANTI-SENSE: No
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: liver
- (ix) FEATURE:
 - (A) NAME/KEY: 5'UTR
 - (B) LOCATION: 1..177
 - (C) IDENTIFICATION METHODS: E
 - (A) NAME/KEY: leader peptide
 - (B) LOCATION: 178..285
 - (C) IDENTIFICATION METHODS: S
 - (A) NAME/KEY: mat peptide
 - (B) LOCATION: 286..756
 - (C) IDENTIFICATION METHODS: S
 - (A) NAME/KEY: 3'UTR
 - (B) LOCATION: 757..1120
 - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

GCCTGGACAG TCAGCAAGGA ATTGTCTCCC AGTGCATTTT GCCCTCCTGG CTGCCAACTC	60
TGGCTGCTAA AGCGGCTGCC ACCTGCTGCA GTCTACACAG CTTCGGGAAG AGGAAAGGAA	120
CCTCAGACCT TCCAGATCGC TTCCTCTCGC AACAACTAT TTGTCGCAGG AATAAAG	177
ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA ATG	225
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala Met	
-35 -30 -25	
AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA GCT GAA GAT GAT GAA AAC	273
Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala Glu Asp Asp Glu Asn	
-20 -15 -10 -5	
CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA GTC ATA	321
Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser Val Ile	
1 5 10	
AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT CGG CCT	369
Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn Arg Pro	
15 20 25	
CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA GAT AAT GCA CCC CGG	417
Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp Asn Ala Pro Arg	
30 35 40	
ACC ATA TTT ATT ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG	465
Thr Ile Phe Ile Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met	
45 50 55 60	
GCT GTA ACT ATC TCT GTG AAG TGT GAG AAA ATT TCA AYT CTC TCC TGT	513
Ala Val Thr Ile Ser Val Lys Cys Glu Lys Ile Ser Xaa Leu Ser Cys	
65 70 75	
GAG AAC AAA ATT TCC TTT AAG GAA ATG AAT CCT CCT GAT AAC ATC	561
Glu Asn Lys Ile Ile Ser Phe Lys Glu Met Asn Pro Pro Asp Asn Ile	
80 85 90	
AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA AGT GTC CCA GGA	609

Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg Ser Val Pro Gly
 95 100 105

CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC GAA GGA TAC TTT 657
 His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr Glu Gly Tyr Phe
 110 115 120
 CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC ATT TTG AAA AAA 705
 Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu Ile Leu Lys Lys
 125 130 135 140
 GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT GTT CAA AAC GAA 753
 Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr Val Gln Asn Glu
 145 150 155
 GAC TAGCTATTAA AATTTTCATGC CGGGCGCAGT GGCTCACGCC TGTAATCCCA 806
 Asp
 GCCCTTTGGG AGGCTGAGGC GGGCAGATCA CCAGAGGTCA GGTGTTCAAG ACCAGCCTGA 866
 CCAACATGGT GAAACCTCAT CTCTACTAAA AATACTAAAA ATTAGCTGAG TGTAGTGACG 926
 CATGCCCTCA ATCCCAGCTA CTCAAGAGGC TGAGGCAGGA GAATCACTTG CACTCCGGAG 986
 GTAGAGGTTG TGGTGAGCCG AGATTGCACC ATTGCGCTCT AGCCTGGGCA ACAACAGCAA 1046
 AACTCCATCT CAAAAAATAA AATAAATAA TAAACAAATA AAAAATTCAT AATGTGAAAA 1106
 AAAAAAAAAA AAAA 1120

(2) INFORMATION FOR SEQ ID NO: 3:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 135 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: exon
 - (B) LOCATION: 1..135
 - (C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA TCT AAA TTA TCA 47
 Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
 -5 1 5 10
 GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT GAC CAA GGA AAT 95
 Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile Asp Gln Gly Asn
 15 20 25
 CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT AGA G 135
 Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys Arg Asp
 30 35 40

(2) INFORMATION FOR SEQ ID NO: 4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 134 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta

(ix) · FEATURE:

- (A) NAME/KEY: exon
(B) LOCATION: 1..134
(C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

AT	AAT	GCA	CCC	CGG	ACC	ATA	TTT	ATT	ATA	AGT	ATG	TAT	AAA	GAT	AGC	47
Asp	Asn	Ala	Pro	Arg	Thr	Ile	Phe	Ile	Ile	Ser	Met	Tyr	Lys	Asp	Ser	
40				45					50						55	
CAG	CCT	AGA	GGT	ATG	GCT	GTA	ACT	ATC	TCT	GTG	AAG	TGT	GAG	AAA	ATT	95
Gln	Pro	Arg	Gly	Met	Ala	Val	Thr	Ile	Ser	Val	Lys	Cys	Glu	Lys	Ile	
			60						65					70		
TCA	ACT	CTC	TCC	TGT	GAG	AAC	AAA	ATT	ATT	TCC	TTT	AAG				134
Ser	Thr	Leu	Ser	Cys	Glu	Asn	Lys	Ile	Ile	Ser	Phe	Lys				
			80					85								

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 87 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon
(B) LOCATION: 1..87
(C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

GAATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG 50
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val
-35 -30 -25

GCA ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G 87
Ala Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala
-20 -15 -10

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon
(B) LOCATION: 1..87
(C) IDENTIFICATION METHODS: S

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CT GAA GAT GAT G
Ala Glu Asp Asp Glu
-10

12

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2167 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human
- (F) TISSUE TYPE: placenta

(ix) FEATURE:

- (A) NAME/KEY: exon + 3'UTR
- (B) LOCATION: 1..2167
- (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GAA ATG AAT CCT CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA	48
Glu Met Asn Pro Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile	
85 90 95 100	
TTC TTT CAG AGA AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA	96
Phe Phe Gln Arg Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu	
105 110 115	
TCT TCA TCA TAC GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC	144
Ser Ser Ser Tyr Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp	
120 125 130	
CTT TTT AAA CTC ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT	192
Leu Phe Lys Leu Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser	
135 140 145	
ATA ATG TTC ACT GTT CAA AAC GAA GAC TAGCTAT TAAAATTTC TGCCGGGCGC	246
Ile Met Phe Thr Val Gln Asn Glu Asp	
150 155	
AGTGGCTCAC GCCTGTAATC CCAGCCCTTT GGGAGGCTGA GGCGGGCAGA TCACCAGAGG	306
TCAGGTGTTT AAGACCAGCC TGACCAACAT GGTGAAACCT CATCTCTACT AAAAATACAA	366
AAAATTAGCT GAGTGTAGTG ACCCATGCC TCAATCCCAG CTACTCAAGA GGCTGAGGCA	426
GGAGAATCAC TTGCACTCCG GAGGTGGAGG TTGTGGTGAG CCGAGATTGC ACCATTGCGC	486
TCTAGCCTGG GCAACAACAG CAAAACCTCCA TCTCAAAAAA TAAAATAAAT AAATAAACAA	546
ATAAAAAATT CATAATGTGA ACTGTCTGAA TTTTATGTT TAGAAAGATT ATGAGATTAT	606
TAGTCTATAA TTGTAATGGT GAAATAAAAT AAATACCAGT CTTGAAAAAC ATCATTAAGA	666
AATGAATGAA CTTTCACAAA AGCAAAACAA CAGACTTTCC CTTATTTAAG TGAATAAAAT	726
AAAATAAAAT AAAATAATGT TTAAAAAATT CATAGTTTGA AAACATTCTA CATTGTTAAT	786
TGGCATATTA ATTATACTTA ATATAATTAT TTTTAAATCT TTTGGGTTAT TAGTCCTAAT	846
GACAAAAGAT ATTGATATTT GAACTTTCTA ATTTTAAAGA ATATCGTTAA ACCATCAATA	906
TTTTTATAAG GAGGCCACTT CACTTGACAA ATTTCTGAAT TTCCTCCAAA GTCAGTATAT	966
TTTTAAAATT CAGTTTGATC CTGAATCCAG CAATATATAA AAGGGATTAT ATACTCTGGC	1026
CAACTGACAT TCATCCTAGG AATGCAAAGA TGGTTTAATA TCCTAAAATC AATTAACATA	1086
ACATACTATA TTAATAAAGT ATCAAAACAG TATTCTCATC TTTTTTCTT TTTTCACAAT	1146
TCCTTGTTTA CACTATCATC TCAATAGATG CAGAAAAAGC ATTTGACAAA ATCCAATTCA	1206
TAATAAAAAAT TCTCAAACCT GAAAGAGAAC ATCATAAAGG CATCTATGAA AAACCTACAG	1266
CTAATATCAT ACTTAACGAT GAAAACTGA ATTATTTTAC CCTAAGATCA AGAATAATGC	1326
AAGCATGTCA GCTCTTGCAA CTTCTATTCA ACATTGTACT GGAGGTTCTA GCCAGAGCAA	1386
CCATACAATA AATAAAAAATA AAAGGCACCC AGATTAGAAA GGAAGTCTTT ATTTGCAGAC	1446
AACATGGTTC TTTATGCAGA AAACCGTCAG GAATACACAC ACATGTTAGA ACTAATAAGT	1506
TCAGCAAGGT TGCAGGTTGC AATATCAATA TGCAAAAAATA CATTGAAGGC TGGGCTCAGT	1566
GGAGATGGCA TGTACCTTTT GTCCCAGCTA CTTGGGAGGC TGAGGTAGGA GGATCACTTG	1626
AGGTGAGGAG TTTGAGGCTA TAGTGCAATG TGATCTTGCC TGTGAATAGC CACTGCACTC	1686
GAGCCTAGGC AACAAAGTGA GACCCCGTCT CCAAAAAAAA AAATGGTATA TTGGTATTTT	1746
TGTATATGAA CAATGAATGA TCTGAAAACA AGAAAAATCC ATTCACGATG GTATTAAAAA	1806
AATAAAATAC AAATAAATTT AGCAAAATAA CAGATAAACT TGTACATCGA AAATTTCAAA	1866

GCACTCTGAG	GGAAATTAAA	GATGATCTAA	ATAATTGGAG	AGACACTCTA	TGATCACTGA	1926
TTGGAAAATT	CATTCAATAT	TGTTAAGATA	ACAATTGTCC	CCAAATTGAT	GCATGCATTG	1986
AATTTAGTCT	TCATCAAAAT	TCCAGCAGGG	TTTTTGCAGA	AATTGACAAG	CTGTACCCAA	2046
AATGTATATG	GAAATGAAAA	GACCCAGAAG	AGCAAATAAT	TTTTTAAAAA	CAAAGTTGGA	2106
AAACTTTTAC	TTCCTAATTT	TAAAACTTAC	TATAAACCTA	AAGTTATCAA	GACCATTTAG	2166
T						2167

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1334 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: intron
 - (B) LOCATION: 1..1334
 - (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GTATTTTTTT	TAATTCGCAA	ACATAGAAAT	GACTAGCTAC	TTCTTCCCAT	TCTGTTTTAC	60
TGCTTACATT	GTTCCGTGCT	AGTCCCAATC	CTCAGATGAA	AAGTCACAGG	AGTGACAATA	120
ATTTCACTTA	CAGGAAACTT	TATAAGGCAT	CCACGTTTTT	TAGTTGGGGT	AAAAAATTGG	180
ATACAATAAG	ACATTGCTAG	GGGTCATGCC	TCTCTGAGCC	TGCCTTTGAA	TCACCAATCC	240
CTTTATTGTG	ATTGCATTAA	CTGTTTAAAA	CCTCTATAGT	TGGATGCTTA	ATCCCTGCTT	300
GTTACAGCTG	AAAATGCTGA	TAGTTTACCA	GGTGTGGTGG	CATCTATCTG	TAATCCTAGC	360
TACTTGGGAG	GCTCAAGCAG	GAGGATTGCT	TGAGGCCAGG	ACTTTGAGGC	TGTAGTACAC	420
TGTGATCGTA	CCTGTGAATA	GCCACTGCAC	TCCAGCCTGG	GTGATATACA	GACCTTGTCT	480
CTAAAATTAA	AAAAAAAAAA	AAAAAAAAACC	TTAGGAAAGG	AAATTGATCA	AGTCTACTGT	540
GCCTTCCAAA	ACATGAATTC	CAAATATCAA	AGTTAGGCTG	AGTTGAAGCA	GTGAATGTGC	600
ATTCTTTTAA	AATACTGAAT	ACTTACCTTA	ACATATATTT	TAAATATTTT	ATTTAGCATT	660
TAAAAGTTAA	AAACAATCTT	TTAGAATTCA	TATCTTTTAA	ATACTCAAAA	AAGTTGCAGC	720
GTGTGTGTTG	TAATACACAT	TAAACTGTGG	GGTTGTTTGT	TTGTTTGAGA	TGCAGTTTCA	780
CTCTGTCAAC	CAGGCTGAAG	TGCAGTGCAG	TGCAGTGGTG	TGATCTCGGC	TCACTACAAC	840
CTCCACCTCC	CACGTTCAAG	CGATTCTCAT	GCCTCAGTCT	CCCGAGTAGG	TGGGATTACA	900
GGCATGCACC	ACTTACACCC	GGCTAATTTT	TGTATTTTTA	GTAGAGCTGG	GGTTTCACCA	960
TGTTGGCCAG	GCTGGTCTCA	AACCCCTAAC	CTCAAGTGAT	CTGCCTGCCT	CAGCCTCCCA	1020
AACAAACAAA	CAACCCACA	GTTTAATATG	TGTTACAACA	CACATGCTGC	AACTTTTATG	1080
AGTATTTTAA	TGATATAGAT	TATAAAAGGT	TGTTTTTAAC	TTTTAAATGC	TGGGATTACA	1140
GGCATGAGCC	ACTGTGCCAG	GCCTGAACTG	TGTTTTTAAA	AATGTCTGAC	CAGCTGTACA	1200
TAGTCTCCTG	CAGACTGGCC	AAGTCTCAAA	GTGGGAACAG	GTGTATTAAG	GACTATCCTT	1260
TGGTTAAATT	TCCGCAATG	TTCCTGTGCA	AGAATTCTTC	TAAC TAGAGT	TCTCATTTAT	1320
TATATTTATT	TCAG					1334

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 4773 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:

(A) NAME/KEY: intron
 (B) LOCATION: 1..4773
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GTAAGACTGA	GCCTTACTTT	GTTTTCAATC	ATGTTAATAT	AATCAATATA	ATTAGAAATA	60
TAACATTATT	TCTAATGTTA	ATATAAGTAA	TGTAATTAGA	AAACTCAAAT	ATCCTCAGAC	120
CAACCTTTTG	TCTAGAACAG	AAATAACAAG	AAGCAGAGAA	CCATTAAAGT	GAATACTTAC	180
TAAAAATTAT	CAAACCTCTT	ACCTATTGTG	ATAATGATGG	TTTTTCTGAG	CCTGTCACAG	240
GGGAAGAGGA	GATACAACAC	TTGTTTTATG	ACCTGCATCT	CCTGAACAAT	CAGTCTTTAT	300
ACAAATAATA	ATGTAGAATA	CATATGTGAG	TTATACATTT	AAGAATAACA	TGTGACTTTC	360
CAGAATGAGT	TCTGCTATGA	AGAATGAAGC	TAATTATCCT	TCTATATTTT	TACACCTTTG	420
TAAATTATGA	TAATATTTTA	ATCCCTAGTT	GTTTTGTTGC	TGATCCTTAG	CCTAAGTCTT	480
AGACACAAGC	TTCAGCTTCC	AGTTGATGTA	TGTTATTTT	AATGTTAATC	TAATTGAATA	540
AAAGTTATGA	GATCAGCTGT	AAAAGTAATG	CTATAATTAT	CTTCAAGCCA	GGTATAAAGT	600
ATTTCTGGCC	TCTACTTTTT	CTCTATTATT	CTCCATTATT	ATTCTCTATT	ATTTTTCTCT	660
ATTTCTCCA	TTATTGTTAG	ATAAACCACA	ATTAACATA	GCTACAGACT	GAGCCAGTAA	720
GAGTAGCCAG	GGATGCTTAC	AAATTGGCAA	TGCTTCAGAG	GAGAATTCCA	TGTCATGAAG	780
ACTCTTTTGG	AGTGGAGATT	TGCCAATAAA	TATCCGCTTT	CATGCCACC	CAGTCCCCAC	840
TGAAAGACAG	TTAGGATATG	ACCTTAGTGA	AGGTACCAAG	GGGCAACTTG	GTAGGGAGAA	900
AAAAGCCACT	CTAAAATATA	ATCCAAGTAA	GAACAGTGCA	TATGCAACAG	ATACAGCCCC	960
CAGACAAATC	CCTCAGCTAT	CTCCCTCCAA	CCAGAGTGCC	ACCCCTTCAG	GTGACAATTT	1020
GGAGTCCCCA	TTCTAGACCT	GACAGGCAGC	TTAGTTATCA	AAATAGCATA	AGAGGCCTGG	1080
GATGGAAGGG	TAGGGTGGAA	AGGGTTAAGC	TATCGGTTAC	TGAACAACAT	AATTAGAAGG	1140
GAAGGAGATG	GCCAAGCTCA	AGCTATGTGG	GATAGAGGAA	AACTCAGCTG	CAGAGGCAGA	1200
TTCAGAAACT	GGGATAAGTC	CGAACCTACA	GGTGGATTCT	TGTTGAGGGA	GACTGGTGAA	1260
AATGTTAAGA	AGATGGAAAT	AATGCTTGGC	ACTTAGTAGG	AACTGGGCAA	ATCCATATTT	1320
GGGGGAGCCT	GAAGTTTATT	CAATTTTGAT	GGCCCTTTTA	AATAAAAAGA	ATGTGGCTGG	1380
GCCTGGTGGC	TCACACCTGT	AATCCCAGCA	CTTTGGGAGG	CCGAGGGGGG	CGGATCACCT	1440
GAAGTCAGGA	GTTCAAGACC	AGCCTGACCA	ACATGGAGAA	ACCCCATCTC	TACTAAAAAT	1500
ACAAATTAG	CTGGGCGTGG	TGGCATATGC	CTGTAATCCC	AGCTACTCGG	GAGGCTGAGG	1560
CAGGAGAATC	TTTTGAACCC	GGGAGGCAGA	GGTTGCGATG	AGCCTAGATC	GTGCCATTGC	1620
ACTCCAGCCT	GGGCAACAAG	AGCAAAACTC	GGTCTCAAAA	AAAAAAGATA	AAAAAGTAAA	1680
TTAACCAAAG	GCATTAGCTT	AATAAATTAA	TACTGTTTTT	AAGTAGGGCG	GGGGGTGGCT	1740
GGAAGAGATC	TGTGTAAATG	AGGGAATCTG	ACATTTAAGC	TTCATCAGCA	TCATAGCAAA	1800
TCTGCTTCTG	GAAGGAACCT	AATAAATATT	AGTTGGAGGG	GGGGAGAGAG	TGAGGGGTGG	1860
ACTAGGACCA	GTTTTAGCCC	TTGTCTTTAA	TCCCTTTTCC	TGCCACTAAT	AAGGATCTTA	1920
GCAGTGTTTA	TAAAGTGTC	CTAGGTTCTA	GATAATAAGA	TACAACAGGC	CAGGCACAGT	1980
GGCTCATGCC	TATAATCCCA	GCACTTTGGG	AGGGCAAGGC	GAGTGTCTCA	CTTGAGATCA	2040
GGAGTTCAAG	ACCAGCCTGG	CCAGCATGGC	GATACTCTGT	CTCTACTAAA	AAAAATACAA	2100
AAATTAGCCA	GGCATGGTGG	CATGCACCTG	TAATCCCAGC	TACTCGTGAG	CCTGAGGCAG	2160
AAGAATCGCT	TGAAACCAGG	AGGTGTAGGC	TGCAGTGAGC	TGAGATCGCA	CCACTGCATC	2220
CCAGCCTGGG	CGACAGAATG	AGACTTTGTC	TCAAAAAAAG	AAAAAGATAC	AACAGGCTAC	2280
CCTTATGTGC	TCACCTTTCA	CTGTTGATTA	CTAGCTATAA	AGTCCTATAA	AGTTCTTTGG	2340
TCAAGAACCT	TGACAACACT	AAGAGGGATT	TGCTTTGAGA	GGTTACTGTC	AGAGTCTGTT	2400
TCATATATAT	ACATATACAT	GTATATATGT	ATCTATATCC	AGGCTTGGCC	AGGGTTCCCT	2460
CAGACTTTCC	AGTGCACTTG	GGAGATGTTA	GGTCAATATC	AACTTTCCCT	GGATTTCAGT	2520
TCAACCCCTT	CTGATGTAAA	AAAAAAGAAA	AAAAAGAAA	AAATCCCTTT	CCCTTGGAG	2580
CACTCAAGTT	TCACCAGGTG	GGGCTTTCCA	AGTTGGGGGT	TCTCCAAGGT	CATTGGGATT	2640
GCTTTTACAT	CCATTTGCTA	TGTACCTTCC	CTATGATGGC	TGGGAGTGGT	CAACATCAAA	2700
ACTAGGAAAG	CTACTGCCCA	AGGATGTCCT	TACCTCTATT	CTGAAATGTG	CAATAAGTGT	2760
GATTAAAGAG	ATTGCCTGTT	CTACCTATCC	ACACTCTCGC	TTTCAACTGT	CACTTTCTTT	2820
TTTTCTTTTT	TTCTTTTTTT	CTTTTTTTTT	GAAACGGAGT	CTCGCTCTGT	CGCCGAGGCT	2880
AGAGTGCAGT	GGCACGATCT	CAGCTCACTG	CAAGCTCTGC	CTCCCGGGTT	CACGCCATTC	2940
TCCTGCCTCA	CCCTCCCAAG	CAGCTGGGAC	TACAGGCGCC	TGCCACCATG	CCCAGCTAAT	3000
TTTTTGATAT	TTTAGTAGAG	ACGGGGTTTC	ACCGTGTTAG	CCAGGATGGT	CTCGATCTCC	3060
TGAACCTGTG	ATCCGCCCCG	CTCAGCCTCC	CAAAGTGCTG	GGATTACAGG	TGTGAGCCAT	3120
CGACCCGGC	TCAAAGTAA	CTTTCTATAC	TGGTTCATCT	TCCCCTGTAA	TGTTACTAGA	3180
GCTTTTGAAG	TTTTGGCTAT	GGATTATTTT	TCATTTATAC	ATTAGATTTT	AGATTAGTTC	3240
CAAATTGATG	CCCACAGCTT	AGGGTCTCTT	CCTAAATTGT	ATATTGTAGA	CAGCTGCAGA	3300
AGTGGGTGCC	AATAGGGGAA	CTAGTTTATA	CTTTCATCAA	CTTAGGACCC	ACACTTGTTG	3360
ATAAAGAACA	AAGGTCAAGA	GTTATGACTA	CTGATTCCAC	AACTGATTGA	GAAGTTGGAG	3420
ATAACCCCGT	GACCTCTGCC	ATCCAGAGTC	TTTCAGGCAT	CTTTGAAGGA	TGAAGAAATG	3480
CTATTTTAAT	TTGGGAGGTT	TCTCTATCAG	TGCTTAGGAT	CATGGGAATC	TGTGCTGCCA	3540
TGAGGCCAAA	ATTAAGTCCA	AAACATCTAC	TGGTTCCAGG	ATTAACATGG	AAGAACCTTA	3600
GGTGGTGCCC	ACATGTTCTG	ATCCATCCTG	CAAAATAGAC	ATGCTGCACT	AACAGGAAAA	3660

```

GTGCAGGCAG CACTACCAGT TGGATAACCT GCAAGATTAT AGTTTCAAGT AATCTAACCA 3720
TTTCTCACAA GGCCCTATTC TGTGACTGAA ACATACAAGA ATCTGCATTT GGCCCTTCTAA 3780
GGCAGGGCCC AGCCAAGGAG ACCATATTCA GGACAGAAAT TCAAGACTAC TATGGAAGTG 3840
GAGTGCTTGG CAGGGAAGAC AGAGTCAAGG ACTGCCAAGT GAGCCAATAC AGCAGGCTTA 3900
CACAGGAACC CAGGGCCTAG CCCTACAACA ATTATTGGGT CTATTCAGTG TAAGTTTTAA 3960
TTTCAGGCTC CACTGAAAGA GTAAGCTAAG ATTCTGGCA CTTTCTGTCT CTCTCACAGT 4020
TGGCTCAGAA ATGAGAACTG GTCAGGCCAG GCATGGTGGC TTACACCTGG AATCCCAGCA 4080
CTTTGGGAGG CCGAAGTGGG AGGGTCACTT GAGGCCAGGA GTTCAGGACC AGCTTAGGCA 4140
ACAAAGTGAG ATACCCCCTG ACCCCTTCTC TACAAAAATA AATTTTAAAA ATTAGCCAAA 4200
TGTGGTGGTG TATACTTACA GTCCCAGCTA CTCAGGAGGC TGAGGCAGGG GGATTGCTTG 4260
AGCCCAGGAA TTCAAGGCTG CAGTGAGCTA TGATTTTACC ACTGCACTTC TGGCTGGGCA 4320
ACAGAGCGAG ACCCTGTCTC AAAGCAAAAA GAAAAAGAAA CTAGAACTAG CCTAAGTTTG 4380
TGGGAGGAGG TCATCATCGT CTTTAGCCGT GAATGGTTAT TATAGAGGAC AGAAATTGAC 4440
ATTAGCCCAA AAAGCTTGTG GTCTTTGCTG GAACTCTACT TAATCTTGAG CAAATGTGGA 4500
CACCACTCAA TGGGAGAGGA GAGAAGTAAG CTGTTTGATG TATAGGGGAA AACTAGAGGC 4560
CTGGAAGTGA ATATGCATCC CATGACAGGG AGAATAGGAG ATTCGGAGTT AAGAAGGAGA 4620
GGAGGTCAGT ACTGCTGTTT AGAGATTTTT TTTATGTAAC TCTTGAGAAG CAAAAGTACT 4680
TTTGTCTCTG TTGGTAATAT ACTTCAAAAC AAACCTCATA TATTCAAATT GTTCATGTCC 4740
TGAAATAATT AGGTAATGTT TTTTCTCTA TAG 4773

```

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8835 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: intron
 - (B) LOCATION: 1..8835
 - (C) IDENTIFICATION METHODS: E
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

```

GTAAGAAATA TCATTCCCTCT TTATTTGGAA AGTCAGCCAT GGCAATTAGA GGTAAATAAG 60
CTAGAAAGCA ATTGAGAGGA ATATAAACCA TCTAGCATCA CTACGATGAG CAGTCAGTAT 120
CAACATAAGA AATATAAGCA AAGTCAGAGT AGAATTTT TCTTTTATCA GATATGGGAG 180
AGTATCACTT TAGAGGAGAG GTTCTCAAAC TTTTGTCTCT CATGTTCCCT TTACACTAAG 240
CACATCACAT GTTAGCATAA GTAACATTTT TAATTAAAAA TAACTATGTA CTTTTTTAAC 300
AACAAAAAAA AGCATAAAGA GTGACACTTT TTTATTTTTA CAAGTGTTTT AACTGGTTTA 360
ATAGAAGCCA TATAGATCTG CTGGATTCTC ATCTGCTTTG CATTGAGACT ACTGCAATAT 420
TGCACAGAAT GCAGCCTCTG GTAAACTCTG TTGTACACTC ATGAGAGAAT GGGTGAAAAA 480
GACAAATTAC GTCTTAGAAT TATTAGAAAT AGCTTTCACT TTAGGAACTC CCTGAGAATT 540
GCTGCTTTAG AGTGGTAAGA TAAATAAGCT TCTCTTTAAA CGGAATCTCA AGACAGAATC 600
AGTTACATTA AAAGCAAACA AAAAATTTGC CCATGGTTAG TCATCTTGTTG AAATCTGCCA 660
CACCTTTGGA CTGGGCTACA ATTGGATAAT ATAGCATTC CCGAGATAAT TTTCTCTCAC 720
AATTAAGGAA AGGGCTGAAT AAATATCTCT GTTTGAAGTT GAATAACAAA AATTAGGACC 780
CCCTAAATTT TAGGGCTCCT GAAATTCGTC TTTTGCCTA TATTCAGCTA CTTTACGTTT 840
TATTAAATCT TCTTTCAGGC CAGGTGCACT AGCTCATGCC TAGAATCTCA GGCAGGCCTG 900
AGCCCAGGAA TTTGAGACCA GCCAGGGCAA CACAGTCTCT ACAAAAAAAT AAAAAATTAC 960
CTGGGTGTGT TGGTGCATGC CTGTAGAACT ACTCAGGATG CTGAGGACTG CTTGAGCCCA 1020
GGATAGCCAA ATCTGTGGTG AGTTCAGCCA CTAACAGAG CGAGACTTTC TCAAAAAAAC 1080
AAACAAAAAA ACAACAAC TTCCTTCAAA ATAACTTTTT ATCTGCAATG TTTTCCTATT 1140
GCCTGTGAGA TTAAATTTAC TCTTTTACCT GATTTCCAAA GCCCTCCATA ATCTAATCCG 1200
ACTTTACCTT GTGTTCACTG CAAAATAGCA GGACTGTTCC ACTACAATCC AAAAATCACA 1260
GGTTGGGTGC AGTGGCTCAC TCCTGTAATC CCAACACTTT GGAAGGCCAA GGCAGGTGGA 1320
TTGCTTCAGC TCAGGAGTTC AAGACCAGCC TGGGCAACAT GGCAAAAAACC CTGTCTCTCC 1380
AAAACATACA AAAATTAGCC AGATGTGTTA GTATGTGCCT GTAGTCCCAA CTAATCAAAA 1440
GGCTAAGGCA AGAGGATCAC TTGAGCCCAG GAGGTCAAGG CTACAGTGAG CCATGTTTAC 1500
TGTGTCAGTG CACTCCAGCC TGGGTGATAG AGCAAGACCA TGTCTCAAAA AAAAAAATAA 1560

```


GAAAAAGAAA	GAAAAAACA	TCGCTCTATT	CAGTTCACCC	CCACCACAAC	ATTGTTTTGA	1620
TTATCACATA	AATGCTGGTC	CATTGCCTTC	TCTATCTATT	CAATCTTTA	AGCATTCCTT	1680
GAGATTCAAC	TCAATTCTCC	TTTTCAAAC	AGGCCATTTA	AACTACATCA	GTTCCATTTT	1740
GATTTTCTTG	CTTTGAGTCT	ACAGACTCAA	AAACAAAAAC	TTAAAAACTT	ATTTTTTAAG	1800
TTTTCTGCTA	CTCTCACTTC	TTCAACACTC	ACATACACGC	ATTCATAATA	AGATGGCAGA	1860
ATGTTCAAGG	ATAAAATGAT	TTATAGAACT	GAAAAGTTAG	GTTTTGATCT	TGTTGCTGTC	1920
AAGATGACTA	CCTACCTGAT	CTCAGGTAAT	TAATTATGTA	GCATGCTCCC	TCATTTTCATC	1980
CCATACCTAT	TCAACAGGAT	TGGAATTCCA	CAGCAAGGAT	AAACATAATC	ATAGTTGCTT	2040
TTCAAGTTCA	AGGCATTTTA	ACTTTTAATC	TAGTAGTATG	TTTGTGTGTG	TGTTGTGTTG	2100
TTGAGATGGA	GCCCTGCTGT	GTCACCCAGG	CTGGAGTGCA	GTGGCACGAA	CTCGGCTCAC	2160
TGCAACCTCT	GCCTCATGGG	TTCAATCAGT	TATTCTGCCT	CAGTGTCCCA	AGTAGCTGGG	2220
ACTACAAGGC	ACATGCCACC	ATGCCTGGCT	AATTTTTTGT	TTTTTAGTAG	AAACAGGGCT	2280
TCACCATTGT	GGCCAGGCTG	GTCTCGAATC	CCTGACCTCA	AGTGATCCAG	CCGCCTCGGC	2340
CTCCCAAAGT	GCTGGGATTA	CAGGCATAAG	CCACCGTGCC	CAGCCTAATA	GTATGTTTTT	2400
AAACTCTTAG	TGGCTTAACA	ATGCTGGTTG	TATAATAAAT	ATGCCATAAA	TATTTACTGT	2460
CTTAGAATTA	TGAAGAAGTG	GTTACTAGGC	CGTTTGCCAC	ATATCAATGG	TTCTCTCCTT	2520
ACAGCTTTAA	TTAGAGTCTA	GAATTGCAGG	TTGGTAGAGC	TGGAACAGAC	CTTAAAGATT	2580
GACTAGCCAA	CTTCCTTGTC	CAATGAGGG	AAC TGAGACC	CTTAAATTA	AGTGACTTGC	2640
CCCAGACAAA	ACTGGAACCTC	ATGTGTCCTA	ATTTCCATCA	TGAAATTCTA	CCATTCACCTA	2700
GCCTCTGGCT	AGTTGTCAAA	GTATTGCATA	ACTAAATTTT	TATGTCTGTT	TTAAAGAACA	2760
AATTGTCACT	GCTTACTCCT	GGGAGGGTCT	TTCTGAGGTG	GTTTATAACT	CTTAAAAAAA	2820
AAAAAGTCAG	TAGTCTGAGA	ATTTTAGACG	AAATAGTCAA	AGCATTTTTA	TCCAATGGAT	2880
CTATAATTTT	CATAGATTAG	AGTTAAATCA	AAGAAACACG	GATGAGAAAG	GAAGAGGAAA	2940
ATTGAGGAGA	GGAGGAATGG	GGATGAGAAC	ACACTACTTG	TAATCAGTCA	TAGATTGACT	3000
GAGAATAAC	AAGAAGAATT	GTAAGAAAAT	AAGAATGAAG	AATTCAAAAT	CAACACATGA	3060
AATAAAAAAG	AACTACTAGG	GAAAAATGGA	GAAGACATTA	GAAAAATTAT	TCTATTTTTA	3120
AAATTCTGTT	TTCAAGGCTTC	CCTCCTGTTC	TTCTCCTTTC	TCATTGGTTT	TCAGGTGGAG	3180
GGAAAGTTTA	AGATGGAAAA	AATATATATA	TTCTACACAT	CCCTTTCTAC	GCTGTGTGCA	3240
TGGCAACAAG	GTTTATCATA	GCAAACTTTT	ATTCATACAA	CATTTATTGA	GTTCTTACTG	3300
TGTGGTAAGC	TCTTTCCAGG	TGTTGAAAAT	TCAGGGGAAA	AAAGACAAC	CATTGTCTTA	3360
AAACTCAGAT	GAAAGCTGAA	CAGACCTATT	TTTAATCAAA	GTAACTCTCA	TTTAGGGTAG	3420
TAAGAGCTAT	TTAAGAAGCA	TGAACAGGTG	TGAAGGAGGT	AGGACTCTGA	GGAGAGAATA	3480
GTTAGCTAGG	AATGAAAGAG	CAGAGAAGTT	TTCTTAGAGG	AACTATTAAA	GCTGGGAGTT	3540
ACGGGATGAA	AGATGAGGCA	GGGTTTGAGC	GCAAAAAAAA	AAAAAAGGCA	GGGGAAGGGG	3600
AAGTTCTGGC	CTGGCAGAGA	GAATAACTGT	GGCAACAATG	GAGGAGAGTC	TGGAAGCAAG	3660
AAAACCAAGT	AGAAGAGTAT	TAAATAGAAA	GATGCCAGGG	GTAAAGAGGG	CTTGATTTAA	3720
AACAGTGCTG	TTGGAGATGG	AGAGGAGATA	CCAAATCTCG	GAGACATTTT	TGAGTTAGAA	3780
CCTACAGTAT	TTATCAGACA	AGGGAAAGAT	TAGACAAAGG	AGTTAAGAA	GACTCCCAGG	3840
TTTCAGTTTG	GGGCAGGTAA	CTAGGACATG	TTTTGAAAA	TAATGTATTG	GATCTCTTAC	3900
CATTGGAAT	ATGTATGTGG	AGCCAAATTA	AAATTTGTAC	ATGTATATAA	CTCTCCCCC	3960
ACCACCAGTA	ACTACTTCCC	TAACCTCTCTA	CTTTGTAGCC	AGACTTCCTA	AAAGAATAGT	4020
TTGTAGTCAC	TGTCTTTACT	TTTCCCCTCT	CATTCTGTCC	TAGATATTTG	TCCACCTACC	4080
ATCTGCTGCG	TCCACTTTAC	CCAAACTGTT	CTACGGTTGC	CCAAAACTTC	CTAATTCATT	4140
AATTCAATGA	ACAAGTTTAA	GCTTATATGT	AAATTAGGAG	CTCTACAGTT	TGATTTTCGAG	4200
CAGCCCCTCC	TGAAACCCTT	TCTCTTTTGA	CTTCTGTGAC	ACATCTCAGA	TTTACAAAAC	4260
TGAACTAATT	ATTTTACACT	TGAGCTGTAT	TTTCGTTCTT	CTTTCTTGAT	GAATGAGGTA	4320
ACCACTCAAC	AAATTGCCCA	AGCCAAAAAC	TACGAAGTCA	TCCTCAGTTC	CTCCTTCTTC	4380
TGTTTGACCC	ACAACAGATC	AGCTGAGAAA	TCCCGTGTGT	TAGTATCTCT	TGAATTCATT	4440
ACCTTAATTT	ATAGCCTCAT	CAACTCTTAA	TTGTTAAAT	TACTTCAGTA	GTTGTTGTCT	4500
GACCTCTGTC	CAATCTTGTT	CAATCAGGTC	CATTCTTTTG	TTCTTGGTGG	TGGTGGTGGT	4560
GTTGACAGAG	TTTCGCTTTT	GCTGCCCAGG	CTGAAGTGCA	GTGGAGCACT	TCAGTCAAC	4620
CACAGCCTCC	TGGGTTTAA	CAGTTCACCC	TCCCGAGTAG	CTGGGACTAC	AGGTATGTGC	4680
CACCACACCC	AGCTAATTTT	GTGTTTTCAG	TAGAGACAGG	GTTTCACCAT	GTTGGTCAGG	4740
CTGGTCTCAA	ACTCCTGACC	TCAAGCAATC	CACCCACCTC	AGCCTCCCAA	AGTGCTGGGA	4800
TTACAGGCAT	GAGCCACTGC	ACACGGACCA	GATCCATTGT	TTATGTTGCT	TCTAGAGTGA	4860
GTTTTTAAAA	CACAAATTTG	ACCATATCTT	TCTCCAATTT	AAGTCAGTAT	TTTTTTTTTT	4920
AGGAAAAAAC	AGTTCAAAC	CTTTAGTCTG	CTTACACAAG	GCCTTTGTAG	TCTGACTCTT	4980
CTTTCCAAGC	TTTTCATCAA	GTATACTGCA	AGTTACATTT	TATGTGAATT	GAATTAGGCA	5040
ACGGTATAAA	AATTATAGTT	TATATGGGCA	AAATGGAAAT	AATGTAACT	CTTCCAAATA	5100
GTTTATCTAG	AATGACATAA	TTTCAAAGCT	GTCAGGTCAA	ATGAGTTATA	AACTGTTAAC	5160
ACTATTGCCA	CATGCAAGTG	TCTCTTATAC	TTGGTAGAAT	TATCTGCTTC	CATGTCATTA	5220
TTATGTAAAT	TAGACTTTAA	ATAACTCAGA	AGTTCTTCAG	ACATACAGGT	TATTATTGTG	5280
CTTTTAAAC	ATAATTTTAA	ATAATTTTAT	ATATGATAAT	GTATCCAAAG	TGCTAAGGGA	5340
TGTATTGTTA	CTGCTGTGCA	AAAAAAAAC	AAAAAAAAC	TCCAAATAAA	TATGTTGAAA	5400
CCAAGTTTAT	ATGCAAGAAA	ACAATATTAA	AAAGGCCAAA	GTACCACCAT	AATAGGCTGT	5460
GTGGAGACGG	CAGGCTACAA	AACACTAGTA	ATAATGCTGA	GAAAGTTGAA	AAAAGAAAGA	5520
AAGCAACAAT	ATGCTTTGGT	TGTTGTAGGT	TTATGTACTC	CAAGAATATC	TCCTCTCAAA	5580

```

CTTTTACGTT TTTTCCAAAG AAAAGTTAAC TTTGGCTGGG CGCAGTGGCT CTTGCCTGTA 5640
GTCCCAGCCT TTGGGAGGCC AAGGCGGGCA GATCACCTGA GGTCAGGAGT TTGAGACCAG 5700
CCTGACCAAA AATGGAGAAA CCCGCCCCCTC TACTACTAA AAGAATACAA AATTAGGCCG 5760
GGCAGAGTGG CTTACCCCTG TGATCCCAGC ACTTTGGGAG GCCGAAGCAG GAAGATCACC 5820
TGAGGTCAGG AGTTCGAGAC CAGCCATGGA GAAACCCGTC TCTACTAAAA ATACAAAATT 5880
AGCCGGGCGT GGTGGTGCAT GACTGTAATC CCAGTACTC AGGAGGCTAA GGCAGAGAAT 5940
CACTTGAACC CAGGCAGTGG AGGTTGCAGT GAGCCGAGAT CGTGCCATTG CACTCCAGCC 6000
TGGGCAACAA GAGCGAAACT CTGTATCCAA AAAACAAAAG AAAAGAAAAG GTAACCTTGA 6060
ACTATGTGAG ATCTTTAGAA ATGCATTCTT TCTGTAAAAT GTGACTACAT TTGCCTTATT 6120
TATGGTAAAA ATGTTGAGGC TCACAAACAA CCATATTTTC TCGGTCTCCC CGCTGCCTAG 6180
CCTTTGTTCA CATTGCTTCT TCTTGGTGGA AGCTCTTCCT CTGGCCTTGA AAATGCCTGC 6240
TTCTCTTTCA AGGTAGCACA GTCATCACTT TCTGTGGTAA CTTTCTCCAG CACCATCAAA 6300
CAGAAAGAAT GAATCTCTTG TAAATTCAGC TCTTACGTCA TTCATTACAT TATTTTGTA 6360
CTCTTTATAG ATTCTTCTCT CCCACTAGAC TCTGAGTCAC TGGAGAGTAG GAGCCAACTC 6420
TCATTCATGT GTGGTTTGGT CAGCTACTGG CCACATTCCT GATGCATAGT TAATGCTCAA 6480
ACCTTAACTG GTGAATCAGC TCAAATATTG TCCTTCTCTA AATCCATTCA CTCATTGACT 6540
AACTATGTAC TCAAAATAGT AAACACCCAGT AATTTAATCC AATTCCTGCC CATACTGCTT 6600
GGTACATTTT AGGTGAATTA GTTTGATAAA TATGTGTGTA TTACATAATA TTAAAGTATG 6660
TACAGAAGAT CATGCTAATC ATAATTCACA ACTGATAACT AATCAAAACAT AATGCTCTC 6720
AGGTTAACAA ATGCTGCTCT TCTCAGTTAA TGCAGTCATT AACAAACACC TTCTGATGCT 6780
GATAATAGGG CCTTGTTTCA CAATGAAGCC ATAAAGGTGA ATAAAGAACA TGCCCTCGTG 6840
GAGCTCACAG CCTAGTCATT ATTGTTCTGA TTTTAAATAT TAATGTTGGT TTGGGTTTTG 6900
GTGAAAAATG TTTAGACTTA TCTTAGTGAT CTTTTCATCC TTTGCTATAT TATTTTCTC 6960
TAAGAGTCTT CCTTATCCCC TCCTTTAAAA AACTAGGTGA TAATTCATAA TTGTAAATTT 7020
AAATATTATA AATAGCTTAT AAAATTTAAT ATTTATAATA TTTAAATGTT TGATAAATAT 7080
TTAAATTTTA TAATATTTAA ATGTTTATTT AAATTCATTT GTACATCAGT TTTTATTTTA 7140
TTTAAATGTG TTGGCCAGGC ATGGTGGCTG ACACCTATAA TCCAGAACT TTGAGAGGCC 7200
AAGTCAGGCA AACCATTTGA GCTCAGGAGT TTGAGACCAC CCTGGGCAAC CTGGTGAAAC 7260
CCTGCTCTCA CCAAACATAT GAAAACCTAT CTGGGTGTGG TGGCACGCAT CTGTGGTCCC 7320
AGATGGGAGT CCCAGGCTAA GATGGGAGAA TCGCTTGAAC CCAGGTGAGA GGGGTGGGGT 7380
GGATGTTGCA GTGAGCTGAG ATCGTGCCAC TGCCTCCAA CCTGGGTGAC AGAGTGAGAC 7440
TCCATCTCAA AAAAAAAAAA TGTTATCTAA ATAAGATAAA TTAATAACT GTTCGCACTT 7500
AGATGAGCAT AAGGAACATA ACCTAGATAA AACTATCAAA TAAGGCCTGG GTACATGAC 7560
TCATGCTGTG AATCTCAAGC ACTTTGGGAG GCCAAAATTA TACAAAGTTA GTTGATATAA 7620
ACCAACTAAC AACTATTTTG GGGTTAGCTT AATTCAGATT AATTTTTTTT AAAGTGAAGT 7680
TTAAATTCCT GCTTACTCTA CCATACATGC TAGGCCTCAT ATTATGCTAG AAAAATTTTG 7740
AGCACAGATT TATGAATACT CTCCTGCATA CCATTTAATT TTTAAACAAA TTTTAATGCA 7800
GTATATATGT GCCTTTTTAC CAACACATTA AATAATAAGA TCTACTGTGA GGACTAAATT 7860
TCTGTAATTT CAAAGTAGTA ATGAGTTTAA ACCATGTCTC AAGATCTCTG CAATAACTGT 7920
AGCACAAACG AAAATAGGTA TTTCTATTAA TGACAGAGTC ACAAGTACTA CTAATAATAC 7980
TGTGTTTGTG TTCCTGCAAC TAATCATGGG AGGAATGCTA AATTTTCAGAG GTTGGTGAAA 8040
ATACATGTGT ATTTTTTTCC CCATCCAAGT TCACAGATTT CTCACACTGA GAACTCCTAT 8100
TCCATAACAA AATTCTGGAA GCCTGCACAC CGTATTGGAA GAAGGGCAGA AAGGAAAAGC 8160
AAATGGAAGG ATTTAAATTT TTTTCAATC CTGTATCCCT TGATTTTACA GCAAGATTGT 8220
ATTTATGTAT TACTTGTGTT AAAAATATAG TATAATCGAG ACTCCAGATC AAAAATCACC 8280
GCAGCTCAGG GAGAAAGAGG GCCACCAAAAT GCCAGAGCCC TTCAGCCTTC TCCCACCCTG 8340
CCTGTACCCT CAGATGGAAG CACTTTTTTA TCATTGTTTC ACCTTTAGCA TTTTGACAAT 8400
GAAGTCACAA ACCTTCAGCC TCTCACCCAT AGGAACCCAC TGGTTGTAAG AGAAGGATGA 8460
AGCCAGTCCT TCCTAAAGGG CACGATTAGA TGTGTTTATG GCATCCTCAG GTGAAACTAT 8520
ATTTATATTG ACAATATATT TATATTTCTC AAGGAATACT AGAATAATGA TTCAGTTTCA 8580
TACTAGGCCA TTTATCTACC CTTTATAATA TTGTTTAATG AGAAAATGCT TTCTATCTTC 8640
CAAATATCTG ATGATTTGTA AGAGAACACT TAAACATGGG TATTCATAAG CTGAAACTTC 8700
TGGCATTATG TGAATGTCAA GATTGTTTAT CAGTATACTA GGTGATTAAC TGACCACTGA 8760
ACTTGAAGGT AGTATAAAGT AGTAGTAAAA GGTACAATCA TTGTCTCTTA ACAGATGGCT 8820
CTTTGCTTTC ATTAG 8835

```

(2) INFORMATION FOR SEQ ID NO:11:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 1371 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:

(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..1371
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

GTAAGGCTAA	TGCCATAGAA	CAAATACCAG	GTTTCAGATAA	ATCTATTCAA	TTAGAAAAGA	60
TGTTGTGAGG	TGAACATTA	AGTGACTCTT	TGTGTCACCA	AATTTCACTG	TAATATTAAT	120
GGCTCTTAAA	AAAATAGTGG	ACCTCTAGAA	ATTAACCACA	ACATGTCCAA	GGTCTCAGCA	180
CCTTGTCACA	CCACGTGTCC	TGGCACTTTA	ATCAGCAGTA	GCTCACTCTC	CAGTTGGCAG	240
TAAGTGCACA	TCATGAAAAT	CCCAGTTTTT	ATGGGAAAAT	CCCAGTTTTT	ATTGGATTTC	300
CATGGGAAAA	ATCCCAGTAC	AAAACCTGGG	GCATTTCAGGA	AATACAATTT	CCCAAAGCAA	360
ATTGGCAAAT	TATGTAAGAG	ATTCTCTAAA	TTTAGAGTTC	CGTGAATTAC	ACCATTTTAT	420
GTAAATATGT	TTGACAAAGTA	AAAATTGATT	CTTTTTTTTT	TTTTCTGTTG	CCCAGGCTGG	480
AGTGCACTGG	CACAATCTCT	GCTCACTGCA	ACCTCCACCT	CCTGGGTTCA	AGCAATTCTC	540
CTGCCTCAGC	CTTCTGAGTA	GCTGGGACTA	CAGGTGCATC	CCGCCATGCC	TGGCTAATTT	600
TTGGGTATTT	TTACTAGAGA	CAGGGTTTTG	GCATGTTGTC	CAGGCTGGTC	TTGGACTCCT	660
GATCTCAGAT	GATCCTCCTG	GCTCGGGCTC	CCAAAGTGCT	GGGATTACAG	GCATGAACCA	720
CCACACATGG	CCTAAAAATT	GATTCTTATG	ATTAATCTCC	TGTGAACAAT	TTGGCTTCAT	780
TTGAAAGTTT	GCCTTCATTT	GAAACCTTCA	TTTAAAAGCC	TGAGCAACAA	AGTGAGACCC	840
CATCTCTACA	AAAAACTGCA	AAATATCCTG	TGGACACCTC	CTACCTTCTG	TGGAGGCTGA	900
AGCAGGAGGA	TCACTTGAGC	CTAGGAATTT	GAGCCTGCAG	TGAGCTATGA	TCCCACCCCT	960
ACACTCCAGC	CTGCATGACA	GTAGACCCTG	ACACACACAC	ACAAAAAATA	ACCTTCATAA	1020
AAAATTATTA	GTTGACTTTT	CTTAGGTGAC	TTCCGTTTAA	AGCAATAAAT	TTAAAAGTAA	1080
AATCTCTAAT	TTTAGAAAAT	TTATTTTATG	TTACATATTG	AAATTTTAA	ACCCTAGGTT	1140
TAAGTTTAT	GTCTAAATTA	CCTGAGAACA	CACTAAGTCT	GATAAGCTTC	ATTTTATGGG	1200
CCTTTTGGAT	GATTATATAA	TATTCTGATG	AAAGCCAAGA	CAGACCCTTA	AACCATAAAA	1260
ATAGGAGTTC	GAGAAAGAGG	AGTAGCAAAA	GTAAAAGCTA	GAATGAGATT	GAATTCTGAG	1320
TCGAAATACA	AAATTTTACA	TATTCTGTTT	CTCTCTTTTT	CCCCCTCTTA	G	1371

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3383 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human
(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: intron
(B) LOCATION: 1..3383
(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GTAAAGTAGA	AATGAATTTA	TTTTTCTTTG	CAAACCTAAGT	ATCTGCTTGA	GACACATCTA	60
TCTCACCATT	GTCAGCTGAG	GAAAAAATAA	AATGGTTCTC	ATGCTACCAA	TCTGCCTTCA	120
AAGAAATGTG	GACTCAGTAG	CACAGCTTTG	GAATGAAGAT	GATCATAAGA	GATACAAAGA	180
AGAACCTCTA	GCAAAAGATG	CTTCTCTATG	CCTTAAAAAA	TTCTCCAGCT	CTTAGAATCT	240
ACAAAATAGA	CTTTGCCTGT	TTCATTGGTC	CTAAGATTAG	CATGAAGCCA	TGGATTCTGT	300
TGTAGGGGGA	GCGTTGCATA	GGAAAAAGGG	ATTGAAGCAT	TAGAATTGTC	CAAAATCAGT	360
AACACCTCCT	CTCAGAAATG	CTTTGGGAAG	AAGCCTGGAA	GGTTCCGGGT	TGGTGGTGGG	420
GTGGGGCAGA	AAATTCCTGA	AGTAGAGGAG	ATAGGAATGG	GTGGGGCAAG	AAGACCACAT	480
TCAGAGGCCA	AAAGCTGAAA	GAAACCATGG	CATTATGAT	GAATTCAGGG	TAATTCAGAA	540
TGGAAGTAGA	GTAGGAGTAG	GAGACTGGTG	AGAGGAGCTA	GAGTGATAAA	CAGGGTGTAG	600
AGCAAGACGT	TCTCTCACCC	CAAGATGTGA	AATTTGGACT	TTATCTTGGA	GATAATAGGG	660
TTAATTAAGC	ACAATATGTA	TTAGCTAGGG	TAAAGATTAG	TTTGTGTGTA	CAAAGACATC	720

CAAAGATACA	GTAGCTGAAT	AAGATAGAGA	ATTTTCTCT	CAAAGAAAGT	CTAAGTAGGC	780
AGCTCAGAA	GTAGTATGGCT	GGAAGCAACC	TGATGATATT	GGGACCCCCA	ACCTTCTTCA	840
GTCTTGTTAC	CATCATCCCC	TAGTTGTTGA	TCTCACTCAC	ATAGTTGAAA	ATCATCATAC	900
TTCCTGGGTT	CATATCCCAG	TTATCAAGAA	AGGGTCAAGA	GAAGTCAGGC	TCATTCCCTT	960
CAAAGACTCT	AATTGGAAGT	TAAACACATC	AATCCCCCTC	ATATTCCATT	GACTAGAATT	1020
TAATCACATG	GCCACACCAA	GTGCAAGGAA	ATCTGGAAAA	TATAATCTTT	ATTCCAGGTA	1080
GCCATATGAC	TCTTTAAAA	TCAGAAATAA	TATATTTTTA	AAATATCATT	CTGGCTTTGG	1140
TATAAAGAAT	TGATGGTGTG	GGGTGAGGAG	GCCAAAATTA	AGGGTTGAGA	GCCTATTATT	1200
TTAGTTATTA	CAAGAAATGA	TGGTGTCTAG	AATTAAGGTA	GACATAGGGG	AGTGCTGATG	1260
AGGAGCTGTG	AATGGATTTT	AGAAACACTT	GAGAGAATCA	ATAGGACATG	ATTTAGGGTT	1320
GGATTTGGAA	AGGAGAAGAA	AGTAGAAAA	ATGATGCCTA	CATTTTTTCAC	TTAGGCAATT	1380
TGTACCATT	AGTGAAATAG	GGAACACAGG	AGGAAGAGCA	GGTTTTGGTG	TATACAAAGA	1440
GGAGGATGGA	TGACGCATTT	CGTTTTGGAT	CTGAGATGTC	TGTGGAACGT	CCTAGTGGAG	1500
ATGTCCACAA	ACTCTTCTAC	ATGTGGTTCT	GAGTTCAGGA	CACAGATTTG	GGCTGGAGAT	1560
AGAGATATTG	TAGGCTTATA	CATAGAAATG	GCATTTGAAT	CTATAGAGAT	AAAAAGACAC	1620
ATCAGAGGAA	ATGTGTAAAG	TGAGAGAGGA	AAAGCCAAGT	ACTGTGCTGG	GGGGAATACC	1680
TACATTTAAA	GGATGCAGTA	GAAAGAAGCT	AATAAACAA	AGAGAGCAGA	CTAACCAAAA	1740
GGGGAGAAGA	AAAACCAAGA	GAATTCCACC	GACTCCCAGG	AGAGCATTTT	AAGATTGAGG	1800
GGATAGGTGT	TGTGTTGAAT	TTTGACGCTT	TGAGAATCAA	GGGCCAGAAC	ACAGCTTTTA	1860
GATTTAGCAA	CAAGGAGTTT	GGTGATCTCA	GTGAAAGCAG	CTTGATGGTG	AAATGGAGGC	1920
AGAGGCAGAT	TGCAATGAGT	GAAACAGTGA	ATGGGAAGTG	AAGAAATGAT	ACAGATAATT	1980
CTTGCTAAAA	GCTTGGCTGT	TAAAAGGAGG	AGAGAAACAA	GACTAGCTGC	AAAGTGAGAT	2040
TGGGTTGATG	GAGCAGTTTT	AAATCTCAAA	ATAAGAGCT	TTGTGCTTTT	TTGATTATGA	2100
AAATAATGTG	TTAATTGTAA	CTAATTGAGG	CAATGAAAA	AGATAATAAT	ATGAAAGATA	2160
AAAATATAAA	AACCAACCAG	AAATAATGAT	AGCTACCATT	TTGATACAAT	ATTTCTACAC	2220
TCCTTTCTAT	GTATATATAC	AGACACAGAA	ATGCTTATAT	TTTTATTAAA	AGGGATTGTA	2280
CTATACCTAA	GCTGCTTTTT	CTAGTTAGTG	ATATATATGG	ACATCTCTCC	ATGGCAACGA	2340
GTAATTGCAG	TTATATTAAG	TTCATGATAT	TTCACAATAA	GGGCATATCT	TTGCCCTTTT	2400
TATTTAATCA	ATTCTTAATT	GGTGAATGTT	GTTTCCAGT	TTGTTGTTGT	TATTAACAAT	2460
GTTCCCATAA	GCATTCCTGT	ACACCAATGT	TCACACATTT	GTCTGATTTT	TTCTTCAGGA	2520
TAAAACCCAG	GAGGTAGAAT	TGCTGGGTTG	ATAGAAGAGA	AAGGATGATT	GCCAAATTAA	2580
AGCTTCAGTA	GAGGGTACAT	GCCGAGCACA	AATGGGATCA	GCCCTAGATA	CCAGAAATGG	2640
CACTTTCTCA	TTTCCCCTTG	GGACAAAAGG	GAGAGAGGCA	ATAACTGTGC	TGCCAGAGTT	2700
AAATTTGTAC	GTGGAGTAGC	AGGAAATCAT	TTGCTGAAAA	TGAAAACAGA	GATGATGTTG	2760
TAGAGGTCCT	GAAGAGAGCA	AAGAAAATTT	GAAATTGCGG	CTATCAGCTA	TGGAAGAGAG	2820
TGCTGAAGTG	GAAAACAAAA	GAAGTATTGA	CAATTGGTAT	GCTTGTAATG	GCACCGATTT	2880
GAACGCTTGT	GCCATTGTTT	ACCAGCAGCA	CTCAGCAGCC	AAGTTTGGAG	TTTTGTAGCA	2940
GAAAGACAAA	TAAGTTAGGG	ATTTAATATC	CTGGCCAAAT	GGTAGACAAA	ATGAACTCTG	3000
AGATCCAGCT	GACACGGGAA	GGAAGGGGAA	ACGGGAAGAG	GTTAGATAGG	AAATACAAGA	3060
GTCAGGAGAC	TGGAAGATGT	TGTGATATTT	AAGAACACAT	AGAGTTGGAG	TAAAAGTGTA	3120
AGAAAACCTAG	AAGGGTAAGA	GACCGGTCAG	AAAGTAGGCT	ATTGGAAGTT	AACACTTCAG	3180
AGGCAGAGTA	GTTCTGAATG	GTAACAAGAA	ATTGAGTGTG	CCTTTGAGAG	TAGGTTAAAA	3240
AACAATAGGC	AACTTTATTG	TAGCTACTTC	TGGAACAGAA	GATTGTCATT	AATAGTTTTA	3300
GAAAACATAA	ATATATAGCA	TACTTATTTG	TCAATTAACA	AAGAACTAT	GTATTTTTTA	3360
ATGAGATTTA	ATGTTTATTG	TAG				3383

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 11464 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: Genomic DNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: human
 - (F) TISSUE TYPE: placenta
- (ix) FEATURE:
 - (A) NAME/KEY: 5'UTR
 - (B) LOCATION: 1..3
 - (C) IDENTIFICATION METHODS: E
 - (A) NAME/KEY: leader peptide
 - (B) LOCATION: 4..82
 - (C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron
 (B) LOCATION: 83..1453
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 1454..1465
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 1466..4848
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: leader peptide
 (B) LOCATION: 4849..4865
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 4866..4983
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 4984..6317
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 6318..6451
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: intron
 (B) LOCATION: 6452..11224
 (C) IDENTIFICATION METHODS: E
 (A) NAME/KEY: mat peptide
 (B) LOCATION: 11225..11443
 (C) IDENTIFICATION METHODS: S
 (A) NAME/KEY: 3'UTR
 (B) LOCATION: 11444..11464
 (C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

AAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA	48
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala	
-35 -30 -25	
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGG CTAATGCCAT	98
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala	
-20 -15 -10	
AGAACAAATA CCAGGTTTCAG ATAAATCTAT TCAATTAGAA AAGATGTTGT GAGGTGAACT	158
ATTAAGTGAC TCTTTGTGTC ACCAAATTTTCT ACTGTAATAT TAATGGCTCT TAAAAAAATA	218
GTGGACCTCT AGAAATTAAC CACAACATGT CCAAGGTCTC AGCACCTTGT CACACCACGT	278
GTCCTGGCAC TTTAATCAGC AGTAGCTCAC TCTCCAGTTG GCAGTAAGTG CACATCATGA	338
AAATCCCAGT TTTTCATGGGA AAATCCCAGT TTTTCATGGGA TTTCCATGGG AAAAAATCCCA	398
GTACAAAACCT GGGTGCATTC AGGAAATACA ATTTCCCAA GCAAATTGGC AAAATTATGTA	458
AGAGATTCTC TAAATTTAGA GTTCCGTGAA TTACACCATT TTATGTAAAT ATGTTTGACA	518
AGTAAAAATT GATTCTTTTT TTTTTTTTCT GTTGCCAGG CTGGAGTGCA GTGGCACAAT	578
CTCTGCTCAC TGCAACCTCC ACCTCCTGGG TTCAAGCAAT TCTCCTGCCT CAGCCTTCTG	638
AGTAGCTGGG ACTACAGGTG CATCCCAGCA TGCCTGGCTA ATTTTGGGT ATTTTACTA	698
GAGACAGGGT TTTGGCATGT TGTCCAGGCT GGTCTTGGAC TCCTGATCTC AGATGATCCT	758
CCTGGCTCGG GCTCCCAAAG TGCTGGGATT ACAGGCATGA ACCACCACAC ATGGCCTAAA	818
AATTGATTCT TATGATTAAT CTCCTGTGAA CAATTTGGCT TCATTTGAAA GTTTGCCTTC	878
ATTTGAAACC TTCATTTAAA AGCCTGAGCA ACAAAGTGAG ACCCCATCTC TACAAAAAAC	938
TGCAAAATAT CCTGTGGACA CCTCCTACCT TCTGTGGAGG CTGAAGCAGG AGGATCACTT	998
GAGCCTAGGA ATTTGAGCCT GCAGTGAGCT ATGATCCCAC CCCTACACTC CAGCCTGCAT	1058
GACAGTAGAC CCTGACACAC ACACACAAAA AAAAACCTTC ATAAAAAATT ATTAGTTGAC	1118
TTTTCTTAGG TGACTTTCCG TTTAAGCAAT AAATTTAAAA GTAAAAATCTC TAATTTTAGA	1178
AAATTTATTT TTAGTTACAT ATTGAAATTT TTAAACCCTA GGTTTAAGTT TTATGTCTAA	1238
ATTACCTGAG AACACACTAA GTCTGATAAG CTTTATTTTA TGGGCCTTTT GGATGATTAT	1298
ATAATATTCT GATGAAAGCC AAGACAGACC CTTAAACCAT AAAAAATAGGA GTTCGAGAAA	1358
GAGGAGTAGC AAAAGTAAAA GCTAGAATGA GATTGAATTC TGAGTCGAAA TACAAAATTT	1418
TACATATTCT GTTTCTCTCT TTTTCCCCCT CTTAG CT GAA GAT GAT G GTAAA	1470
Ala Glu Asp Asp Glu	
-10	
GTAGAAATGA ATTTATTTTT CTTTGCAAAC TAAGTATCTG CTTGAGACAC ATCTATCTCA	1530
CCATTGTCAG CTGAGGAAAA AAAAAATGG TTCTCATGCT ACCAATCTGC CTTCAAAGAA	1590
ATGTGGACTC AGTAGCACAG CTTTGGAAATG AAGATGATCA TAAGAGATAC AAAGAAGAAC	1650

CTCTAGCAAA	AGATGCTTCT	CTATGCCTTA	AAAAATTCTC	CAGCTCTTAG	AATCTACAAA	1710
ATAGACTTTG	CCTGTTTCAT	TGGTCCTAAG	ATTAGCATGA	AGCCATGGAT	TCTGTTGTAG	1770
GGGGAGCGTT	GCATAGGAAA	AAGGGATTGA	AGCATTAGAA	TTGTCCAAAA	TCAGTAACAC	1830
CTCCTCTCAG	AAATGCTTTG	GGAAGAAGCC	TGGAAGGTTT	CGGGTTGGTG	GTGGGGTGGG	1890
GCAGAAAATT	CTGGAAGTAG	AGGAGATAGG	AATGGGTGGG	GCAAGAAGAC	CACATTCAGA	1950
GGCCAAAAGC	TGAAAGAAAC	CATGGCATT	ATGATGAATT	CAGGGTAATT	CAGAATGGAA	2010
GTAGAGTAGG	AGTAGGAGAC	TGGTGAGAGG	AGCTAGAGTG	ATAAACAGGG	TGTAGAGCAA	2070
GACGTTTCT	CACCCCAAGA	TGTGAAATTT	GGACTTTATC	TTGGAGATAA	TAGGGTTAAT	2130
TAAGCACAAT	ATGTATTAGC	TAGGGTAAAG	ATTAGTTTGT	TGTAACAAAAG	ACATCCAAAG	2190
ATACAGTAGC	TGAATAAGAT	AGAGAATTTT	TCTCTCAAAG	AAAGTCTAAG	TAGGCAGCTC	2250
AGAAGTAGTA	TGGCTGGAAG	CAACCTGATG	ATATTGGGAC	CCCCAACCTT	CTTCAGTCTT	2310
GTACCCATCA	TCCCCTAGTT	GTTGATCTCA	CTCACATAGT	TGAAAATCAT	CATACTTCCT	2370
GGGTTTCATAT	CCCAGTTATC	AAGAAAGGGT	CAAGAGAAGT	CAGGCTCAT	CCTTTCAAAG	2430
ACTCTAATTG	GAAGTTAAAC	ACATCAATCC	CCCTCATATT	CCATTGACTA	GAATTTAATC	2490
ACATGGCCAC	ACCAAGTGCA	AGGAAATCTG	GAAAATATAA	TCTTTATTCC	AGGTAGCCAT	2550
ATGACTCTTT	AAAATTCAGA	AATAATATAT	TTTTAAAATA	TCATTCTGGC	TTTGGTATAA	2610
AGAATTGATG	GTGTGGGGTG	AGGAGGCCAA	AATTAAGGGT	TGAGAGCCTA	TTATTTTAGT	2670
TATTACAAGA	AATGATGGTG	TCATGAATTA	AGGTAGACAT	AGGGGAGTGC	TGATGAGGAG	2730
CTGTGAATGG	ATTTTAGAAA	CACCTGAGAG	AATCAATAGG	ACATGATTTA	GGGTTGGATT	2790
TGGAAGAGGAG	AGAAAGTAG	AAAAGATGAT	GCCTACATTT	TTCACTTAGG	CAATTTGTAC	2850
CATTCAGTGA	AATAGGGAAC	ACAGGAGGAA	GAGCAGGTTT	TGGTGTATAC	AAAGAGGAGG	2910
ATGGATGACG	CATTTCTGTT	TGGATCTGAG	ATGTCTGTGG	AACGTCCTAG	TGGAGATGTC	2970
CACAACTCT	TCTACATGTG	GTTCTGAGTT	CAGGACACAG	ATTTGGGCTG	GAGATAGAGA	3030
TATTGTAGGC	TTATACATAG	AAATGGCATT	TGAATCTATA	GAGATAAAAA	GACACATCAG	3090
AGGAATGTG	TAAAGTGAGA	GAGGAAAAGC	CAAGTACTGT	GCTGGGGGGA	ATACCTACAT	3150
TTAAAGGATG	CAGTAGAAAG	AAGCTAATAA	ACAACAGAGA	GCAGACTAAC	CAAAAGGGGA	3210
GAAGAAAAAC	CAAGAGAATT	CCACCGACTC	CCAGGAGAGC	ATTTCAAGAT	TGAGGGGATA	3270
GGTGTGTGT	TGAATTTTGC	AGCCTTGAGA	ATCAAGGGCC	AGAACACAGC	TTTTAGATTT	3330
AGCAACAAGG	AGTTTGGTGA	TCTCAGTGAA	AGCAGCTTGA	TGGTGAAATG	GAGGCAGAGG	3390
CAGATGTCAA	TGAGTGAAAC	AGTGAATGGG	AAGTGAAGAA	ATGATACAGA	TAATTTCTTC	3450
TAAAAGCTTG	GCTGTTAAAA	GGAGGAGAGA	AACAAGACTA	GCTGCAAAGT	GAGATTGGGT	3510
TGATGGAGCA	GTTTTAAATC	TCAAAATAAA	GAGCTTTGTG	CTTTTTTTGAT	TATGAAAATA	3570
ATGTGTTAAT	TGTAACATAA	TGAGGCAATG	AAAAAAGATA	ATAATATGAA	AGATAAAAAAT	3630
ATAAAAACCA	CCCAGAAATA	ATGATAGCTA	CCATTTTGAT	ACAATATTTT	TACACTCCTT	3690
TCTATGTATA	TATACAGACA	CAGAAATGCT	TATATTTTTA	TTAAAAGGGA	TTGTACTATA	3750
CCTAAGCTGC	TTTTTCTAGT	TAGTGATATA	TATGGACATC	TCTCCATGGC	AACGAGTAAT	3810
TGCAGTTATA	TTAAGTTTAT	GATATTTTAC	AATAAGGGCA	TATCTTTGCC	CTTTTATTTT	3870
AATCAATTCT	TAATTGGTGA	ATGTTTGT	CCAGTTTGT	GTTGTTATTA	ACAATGTTCC	3930
CATAAGCATT	CCTGTACACC	AATGTTTACA	CATTTGTCTG	ATTTTTTCTT	CAGGTATAAAA	3990
CCCAGGAGGT	AGAATTGCTG	GGTTGATAGA	AGAGAAAGGA	TGATTGCCAA	ATTAAAGCTT	4050
CAGTAGAGGG	TACATGCCGA	GCACAAATGG	GATCAGCCCT	AGATACCAGA	AATGGCACTT	4110
TCTCATTTCC	CCTTGGGACA	AAAGGGAGAG	AGGCAATAAC	TGTGCTGCCA	GAGTTAAATT	4170
TGTACGTGGA	GTAGCAGGAA	ATCATTTGCT	GAAAATGAAA	ACAGAGATGA	TGTTGTAGAG	4230
GTCCTGAAGA	GAGCAAAGAA	AATTTGAAAT	TGCGGCTATC	AGCTATGGAA	GAGATGCTG	4290
AACTGGAAAA	CAAAAGAAAT	ATTGACAATT	GGTATGCTTG	TAATGGCACC	GATTTGAACG	4350
CTTGTGCCAT	TGTTTACCAG	CAGCACTCAG	CAGCCAAGTT	TGGAGTTTGT	TAGCAGAAAG	4410
ACAAATAAGT	TAGGGATTTA	ATATCCTGGC	CAAATGGTAG	ACAAAATGAA	CTCTGAGATC	4470
CAGCTGCACA	GGGAAGGAAG	GGAAGACGGG	AAGAGGTTAG	ATAGGAAATA	CAAGAGTCAG	4530
GAGACTGGAA	GATGTTGTGA	TATTTAAGAA	CACATAGAGT	TGGAGTAAAA	GAGTAAGAAA	4590
ACTAGAAGGG	TAAGAGACCG	GTCAGAAAAGT	AGGCTATTTG	AAGTTAACAC	TTCAGAGGCA	4650
GAGTAGTTCT	GAATGGTAAC	AAGAAATTGA	GTGTGCCTTT	GAGAGTAGGT	TAAAAACAA	4710
TAGGCAACTT	TATTGTAGCT	ACTTCTGGAA	CAGAAGATTG	TCATTAATAG	TTTTAGAAAA	4770
CTAAATATA	TAGCATACTT	ATTTGTCAAT	TAACAAAGAA	ACTATGTATT	TTTAAATGAG	4830
ATTTAATGTT	TATTGTAG	AA AAC CTG	GAA TCA GAT	TAC TTT GGC	AAG CTT	4880
		Glu Asn Leu	Glu Ser Asp	Tyr Phe Gly	Lys Leu	
		-5		1	5	
GAA TCT AAA	TTA TCA GTC	ATA AGA AAT	TTG AAT GAC	CAA GTT CTC	TTC	4928
Glu Ser Lys	Leu Ser Val	Ile Arg Asn	Leu Asn Asp	Gln Val Leu	Phe	
	10	15	20			
ATT GAC CAA	GGA AAT CGG	CCT CTA TTT	GAA GAT ATG	ACT GAT TCT	GAC	4976
Ile Asp Gln	Gly Asn Arg	Pro Leu Phe	Glu Asp Met	Thr Asp Ser	Asp	
	25	30	35			
TGT AGA G	GTATTTTTT	TTAATTCGCA	AACATAGAAA	TGACTAGCTA	CTTCTTCCCA	5032
Cys Arg Asp						
	40					
TTCTGTTTTA	CTGCTTACAT	TGTTCCGTGC	TAGTCCCAAT	CCTCAGATGA	AAAGTCACAG	5092
GAGTGACAAT	AATTTCACTT	ACAGGAAACT	TTATAAGGCA	TCCACGTTTT	TTAGTTGGGG	5152

TAAAAAATTG	GATACAATAA	GACATTGCTA	GGGGTCATGC	CTCTCTGAGC	CTGCCTTTGA	5212
ATCACCAATC	CCTTTATTGT	GATTGCATTA	ACTGTTTAAA	ACCTCTATAG	TTGGATGCTT	5272
AATCCCTGCT	TGTTACAGCT	GAAAATGCTG	ATAGTTTACC	AGGTGTGGTG	GCATCTATCT	5332
GTAATCCTAG	CTACTGGGA	GGCTCAAGCA	GGAGGATTGC	TTGAGGCCAG	GACTTTGAGG	5392
CTGTAGTACA	CTGTGATCGT	ACCTGTGAAT	AGCCACTGCA	CTCCAGCCTG	GGTGATATAC	5452
AGACCTTGTC	TCTAAAATTA	AAAAAAAAAA	AAAAAAAAAC	CTTAGGAAAG	GAAATTGATC	5512
AAGTCTACTG	TGCCTTCCAA	AACATGAATT	CCAAATATCA	AAGTTAGGCT	GAGTTGAAGC	5572
AGTGAATGTG	CATTCTTTAA	AAATACTGAA	TACTTACCTT	AACATATATT	TTAAATATTT	5632
TATTTAGCAT	TTAAAAGTTA	AAAACAATCT	TTTGAATTC	ATATCTTTAA	AATACTCAAA	5692
AAAGTTGCAG	CGTGTGTGTT	GTAATACACA	TTAAACTGTG	GGGTGTGTTG	TTTGTGTTGAG	5752
ATGCAGTTTC	ACTCTGTCAC	CCAGGCTGAA	GTGCAGTGCA	GTGCAGTGGT	GTGATCTCGG	5812
CTCACTACAA	CCTCCACCTC	CCACGTTCAA	GCGATTCTCA	TGCCTCAGTC	TCCCCAGTAG	5872
GTGGGATTAC	AGGCATGCAC	CACCTACACC	CGGCTAATTT	TTGTATTTTT	AGTAGAGCTG	5932
GGGTTTCACC	ATGTTGGCCA	GGCTGGTCTC	AAACCCCTAA	CCTCAAGTGA	TCTGCCTGCC	5992
TCAGCCTCCC	AAACAAACAA	ACAACCCAC	AGTTTAATAT	GTGTTACAAC	ACACATGCTG	6052
CAACTTTTAT	GAGTATTTTA	ATGATATAGA	TTATAAAAGG	TTGTTTTTAA	CTTTTAAATG	6112
CTGGGATTAC	AGGCATGAGC	CACTGTGCCA	GGCCTGAACT	GTGTTTTTAA	AAATGTCTGA	6172
CCAGCTGTAC	ATAGTCTCCT	GCAGACTGGC	CAAGTCTCAA	AGTGGGAACA	GGTGATTAA	6232
GGACTATCCT	TTGGTTAAAT	TTCCGCAAAT	GTTCTGTGTC	AAGAATTCTT	CTAACTAGAG	6292
TTCTCATTTA	TTATATTTAT	TTCAG	AT AAT GCA CCC	CGG ACC ATA	TTT ATT	6343
		Asp Asn Ala Pro	Arg Thr Ile Phe Ile			
		40		45		
ATA AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC						6391
Ile Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile						
50	55	60				
TCT GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT						6439
Ser Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile						
65	70	75	80			
ATT TCC TTT AAG GTAAG ACTGAGCCTT ACTTTGTTTT CAATCATGTT AATATAATCA						6496
Ile Ser Phe Lys						
ATATAATTAG AAATATAACA TTATTTCTAA TGTTAATATA AGTAATGTAA TTAGAAAAC						6556
CAAATATCCT CAGACCAACC TTTTGTCTAG AACAGAAATA ACAAGAAGCA GAGAACCATT						6616
AAAGTGAATA CTTACTAAAA ATTATCAAAC TCTTTACCTA TTGTGATAAT GATGGTTTTT						6676
CTGAGCCTGT CACAGGGGAA GAGGAGATAC AACACTTGTT TTATGACCTG CATCTCCTGA						6736
ACAATCAGTC TTTATACAAA TAATAATGTA GAATACATAT GTGAGTTATA CATTTAAGAA						6796
TAACATGTGA CTTTCCAGAA TGAGTTCTGC TATGAAGAAT GAAGCTAATT ATCCTTCTAT						6856
ATTTCTACAC CTTTGTAAT TATGATAATA TTTTAAATCCC TAGTTGTTTT GTTGCTGATC						6916
CTTAGCCTAA GTCTTAGACA CAAGCTTCAG CTTCAGTTG ATGTATGTTA TTTTAAATGT						6976
TAATCTAATT GAATAAAAGT TATGAGATCA CTTGTAAAAG TAATGCTATA ATTATCTTCA						7036
AGCCAGGTAT AAAGTATTTT TGGCCTCTAC TTTTCTCTA TTATTCTCCA TTATTATTCT						7096
CTATTATTTT TCTCTATTTT CTCCATTATT GTTAGATAAA CCACAATTAA CTATAGCTAC						7156
AGACTGAGCC AGTAAGAGTA GCCAGGGATG CTTACAAATT GGCAATGCTT CAGAGGAGAA						7216
TTCCATGTCA TGAAGACTCT TTTTGAGTGG AGATTTGCCA ATAAATATCC GCTTTCATGC						7276
CCACCCAGTC CCACTGAAA GACAGTTAGG ATATGACCTT AGTGAAGGTA CCAAGGGGCA						7336
ACTTGGTAGG GAGAAAAAAG CCACTCTAAA ATATAATCCA AGTAAGAACA GTGCATATGC						7396
AACAGATACA GCCCCAGAC AAATCCCTCA GCTATCTCCC TCCAACCAGA GTGCCACCCC						7456
TTCAGGTGAC AATTTGGAGT CCCCATTCTA GACCTGACAG GCAGCTTAGT TATCAAAATA						7516
GCATAAGAGG CCTGGGATGG AAGGGTAGGG TACCATGCT TAAGCATGCT GTTACTGAAC						7576
AACATAATTA GAAGGGAAGG AGATGGCCAA GCTCAAGCTA TGTGGGATAG AGGAAAAC						7636
AGCTGCAGAG GCAGATTGAG AAAGTGGGAT AAGTCCGAAC CTACAGGTGG ATTCTTGTG						7696
AGGGAGACTG GTGAAAATGT TAAGAAGATG GAAATAATGC TTGGCACTTA GTAGGAAC						7756
GGCAAAATCCA TATTTGGGGG AGCCTGAAGT TTATTTCAATT TTGATGGCCC TTTTAAATAA						7816
AAAGAATGTG GCTGGGCGTG GTGGCTCACA CCTGTAATCC CAGCACTTTG GGAGGCCGAG						7876
GGGGCGGGAT CACCTGAAGT CAGGAGTTCA AGACCAGCCT GACCAACATG GAGAAACCCC						7936
ATCTCTACTA AAAATACAAA ATTAGCTGGG CGTGGTGGCA TATGCCTGTA ATCCCAGCTA						7996
CTCGGGAGGC TGAGGCAGGA GAATCTTTTG AACC CGGGAG GCAGAGGTTG CGATGAGCCT						8056
AGATCGTGCC ATTGCACTCC AGCCTGGGCA ACAAGAGCAA AACTCGGTCT CAAAAA						8116
AAAAAAAAG TGAAATTAAC CAAAGGCATT AGCTTAATAA TTTAATACTG TTTTAAAGTA						8176
GGGCGGGGGG TGGCTGGAAG AGATCTGTGT AAATGAGGGA ATCTGACATT TAAGCTTCAT						8236
CAGCATCATA GCAAATCTGC TTCTGGAAGG AACTCAATAA ATATTAGTTG GAGGGGGGGA						8296
GAGAGTGAGG GGTGGACTAG GACCAGTTTT AGCCCTTGTC TTTAATCCCT TTTCTGCCA						8356
CTAATAAGGA TCTTAGCAGT GGTATAAAA GTGGCCTAGG TTCTAGATAA TAAGATACAA						8416
CAGGCCAGGC ACAGTGGCTC ATGCCATATA TCCAGCACT TTGGGAGGGC AAGGCGAGTG						8476
TCTCACTTGA GATCAGGAGT TCAAGACCAG CCTGGCCAGC ATGGCGATAC TCTGTCTCTA						8536
CTAAAAAAA TACAAAAATT AGCCAGGCAT GGTGGCATGC ACCTGTAATC CCAGCTACTC						8596
GTGAGCCTGA GGCAGAAGAA TCGCTTGAAG CCAGGAGGTG TAGGCTGCAG TGAGCTGAGA						8656
TCGCACCACT GCACTCCAGC CTGGGCGACA GAATGAGGCT TTGTCTCAAA AAAAGAAAAA						8716

GATACAACAG	GCTACCCTTA	TGTGCTCACC	TTTCACTGTT	GATTACTAGC	TATAAAGTCC	8776
TATAAAGTTC	TTTGGTCAAG	AACCTTGACA	ACACTAAGAG	GGATTTGCTT	TGAGAGGTTA	8836
CTGTCAGAGT	CTGTTTCATA	TATATACATA	TACATGTATA	TATGTATCTA	TATCCAGGCT	8896
TGGCCAGGGT	TCCCTCAGAC	TTTCCAGTGC	ACTTGGGAGA	TGTTAGGTCA	ATATCAACTT	8956
TCCCTGGATT	CAGATTCAAC	CCCTTCTGAT	GTAAAAAATA	AAAAAATAAA	GAAAGAAATC	9016
CCTTTCCCTT	TGGAGCACTC	AAGTTTCACC	AGGTGGGGCT	TTCCAAGTTG	GGGGTTCTCC	9076
AAGGTCATTG	GGATTGCTTT	CACATCCATT	TGCTATGTAC	CTTCCCTATG	ATGGCTGGGA	9136
GTGGTCAACA	TCAAAACTAG	GAAAGCTACT	GCCCAAGGAT	GTCCTTACCT	CTATTCTGAA	9196
ATGTGCAATA	AGTGTGATTA	AAGAGATTGC	CTGTTCTACC	TATCCACACT	CTCGCTTTCA	9256
ACTGTAACCT	TCTTTTTTTC	TTTTTTTCTT	TTTTTCTTTT	TTTTTGAAAC	GGAGTCTCGC	9316
TCTGTCGCCC	AGCTAGAGT	GCAGTGGCAC	GATCTCAGCT	CACTGCAAGC	TCTGCCTCCC	9376
GGGTTCACGC	CATTCTCCTG	CCTCACCTC	CCAAGCAGCT	GGGACTACAG	GCGCCTGCCA	9436
CCATGCCCAG	CTAATTTTTT	GTATTTTTAG	TAGAGACGGG	GTTTCACCGT	GTTAGCCAGG	9496
ATGGTCTCGA	TCTCCTGAAC	TTGTGATCCG	CCCGCTCAG	CCTCCCAAAG	TGCTGGGATT	9556
ACAGGCGTGA	GCCATCGCAC	CCGGCTCAAC	TGTAACTTTC	TATACTGGTT	CATCTTCCCC	9616
TGTAATGTTA	CTAGAGCTTT	TGAAGTTTTG	GCTATGGATT	ATTTCTCATT	TATACATTAG	9676
ATTTTCAGATT	AGTTCCAAAT	TGATGCCAC	AGCTTAGGGT	CTCTTCCTAA	ATTGTATATT	9736
GTAGACAGCT	GCAGAAGTGG	GTGCCAATAG	GGGAACTAGT	TTATACTTTC	ATCAACTTAG	9796
GACCCACACT	TGTTGATAAA	GAACAAAGGT	CAAGAGTTAT	GACTACTGAT	TCCACAACCTG	9856
ATTGAGAAGT	TGGAGATAAC	CCCGTGACCT	CTGCCATCCA	GAGTCTTTCA	GGCATCTTTG	9916
AAGGATGAAG	AAATGCTATT	TTAATTTTGG	AGGTTTCTCT	ATCAGTGCTT	AGGATCATGG	9976
GAATCTGTGC	TGCCATGAGG	CCAAAATTAA	GTCCAAAACA	TCTACTGGTT	CCAGGATTAA	10036
CATGGAAGAA	CCTTAGGTGG	TGCCCACATG	TTCTGATCCA	TCCTGCAAAA	TAGACATGCT	10096
GCCTAACAG	GAAAAGTGCA	GGCAGCACTA	CCAGTTGGAT	AACCTGCAAG	ATTATAGTTT	10156
CAAGTAATCT	AACCATTCTT	CACAAGGCCC	TATTCTGTGA	CTGAAACATA	CAAGAATCTG	10216
CAATTGGCCT	TCTAAGGCAG	GGCCAGCCA	AGGAGACCAT	ATTCAGGACA	GAAATTCAAG	10276
ACTACTATGG	AACTGGAGTG	CTTGGCAGGG	AAGACAGAGT	CAAGGACTGC	CAACTGAGCC	10336
AATACAGCAG	GCTTACACAG	GAACCCAGGG	CCTAGCCCTA	CAACAATTAT	TGGGTCTATT	10396
CACTGTAAGT	TTTAATTTCA	GGCTCCACTG	AAAGAGTAAG	CTAAGATTCC	TGGCACTTTC	10456
TGTCTCTCTC	ACAGTTGGCT	CAGAAATGAG	AACTGGTCAG	GCCAGGCATG	TGGGCTTACA	10516
CCTGGAATCC	CAGCACTTTG	GGAGGCCGAA	GTGGGAGGGT	CACTTGAGGC	CAGGAGTTCA	10576
GGACCAGCTT	AGGCAACAAA	GTGAGATACC	CCCTGACCCC	TTCTCTACAA	AAATAAATTT	10636
TAAAAATTAG	CCAAATGTGG	TGGTGTATAC	TTACAGTCCC	AGCTACTCAG	GAGGCTGAGG	10696
CAGGGGGATT	GCTTGAGCCC	AGGAATTCAA	GGCTGCAGTG	AGCTATGATT	TCACCACTGC	10756
ACTTCTGGCT	GGGCAACAGA	GCGAGACCTT	GTCTCAAAGC	AAAAAGAAAA	AGAACTAGA	10816
ACTAGCCTAA	GTTTGTGGGA	GGAGGTCATC	ATCGTCTTTA	GCCGTGAATG	GTTATTATAG	10876
AGGACAGAAA	TTGACATTAG	CCCAAAAAGC	TTGTGGTCTT	TGCTGGAACT	CTACTTAATC	10936
TTGAGCAAAT	GTGGACACCA	CTCAATGGGA	GAGGAGAGAA	GTAAGCTGTT	TGATGTATAG	10996
GGGAAAACCTA	GAGGCCTGGA	ACTGAATATG	CATCCCATGA	CAGGGAGAAT	AGGAGATTCTG	11056
GAGTTAAGAA	GGAGAGGAGG	TCAGTACTGC	TGTTTCAGAGA	TTTTTTTTTAT	GTAACCTTTG	11116
AGAAGCAAAA	CTACTTTTGT	TCTGTTTGGT	AATATACTTC	AAAACAAACT	TCATATATTC	11176
AAATTGTTCA	TGTCCTGAAA	TAATTAGGTA	ATGTTTTTTT	CTCTATAG	GAA ATG AAT	11233

Glu Met Asn
85

CCT	CCT	GAT	AAC	ATC	AAG	GAT	ACA	AAA	AGT	GAC	ATC	ATA	TTC	TTT	CAG	11281
Pro	Pro	Asp	Asn	Ile	Lys	Asp	Thr	Lys	Ser	Asp	Ile	Ile	Phe	Phe	Glu	
		90					95					100				
AGA	AGT	GTC	CCA	GGA	CAT	GAT	AAT	AAG	ATG	CAA	TTT	GAA	TCT	TCA	TCA	11329
Arg	Ser	Val	Pro	Gly	His	Asp	Asn	Lys	Met	Gln	Phe	Glu	Ser	Ser	Ser	
		105					110					115				
TAC	GAA	GGA	TAC	TTT	CTA	GCT	TGT	GAA	AAA	GAG	AGA	GAC	CTT	TTT	AAA	11377
Tyr	Glu	Gly	Tyr	Phe	Leu	Ala	Cys	Glu	Lys	Glu	Arg	Asp	Leu	Phe	Lys	
		120				125				130					135	
CTC	ATT	TTG	AAA	AAA	GAG	GAT	GAA	TTG	GGG	GAT	AGA	TCT	ATA	ATG	TTC	11425
Leu	Ile	Leu	Lys	Lys	Glu	Asp	Glu	Leu	Gly	Asp	Arg	Ser	Ile	Met	Phe	
			140						145					150		
ACT	GTT	CAA	AAC	GAA	GAC	TAGCTATTAA	AATTTTCATGC	C								11464
Thr	Val	Gln	Asn	Glu	Asp											
			155													

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28994 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Genomic DNA

(vi) ORIGINAL SOURCE:

(A) ORGANISM: human

(F) TISSUE TYPE: placenta

(ix) FEATURE:

(A) NAME/KEY: 5'UTR

(B) LOCATION: 1..15606

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 15607..15685

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 15686..17056

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 17057..17068

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 17069..20451

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: leader peptide

(B) LOCATION: 20452..20468

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: mat peptide

(B) LOCATION: 20469..20586

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 20587..21920

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: mat peptide

(B) LOCATION: 21921..22054

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: intron

(B) LOCATION: 22055..26827

(C) IDENTIFICATION METHODS: E

(A) NAME/KEY: mat peptide

(B) LOCATION: 26828..27046

(C) IDENTIFICATION METHODS: S

(A) NAME/KEY: 3'UTR

(B) LOCATION: 27047..28994

(C) IDENTIFICATION METHODS: E

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

ACTTGCCTTA	AAAGCTTTGC	ATAGGTAGAC	AACATTAGAT	TAATTTTCCTT	GCTCACATCT	60
GTTCAAGAAA	AATCATTTAA	GTTATAAAAT	ATAACAAACC	TTCTGCATTA	TAAGACTGAT	120
GTTTAGAAAT	ATAAACATTT	TATACATCAC	CATTTAAATC	TTTCTCCAAG	GCTTCATCTT	180
TATAAAATAG	TCCGGAAATT	TCAGAGAAAG	ATGAATCTGA	TTTTCCAAGA	GAGGACAGCT	240
GTGGACTATC	TGGCACTGGA	GACTAAATAA	AGAAAGCAGG	TACAGTCAAT	AAGATCTTCA	300
GGACATATAC	ATTTTGTTTA	TTAAGAAAAA	GCAAATAAAA	CATTTTTCAG	AAAAAGGCAA	360
ACATGCTAGA	AAGCATATGA	CTTAGTCATT	TGAGTTTTTA	TTATTAAGGA	AATTTACAGG	420
CCCAAGAAAC	ACCTTGCTCA	ATATATTAAA	TTTTATTTTG	GTTTCAACT	AGACTTTGCT	480
TTTCATTTGT	TTGTTTTTGT	GACAAGTTCT	CGCTCTGTCA	CCTAGGCCAA	AGTGTAGTGA	540
CACAATCTTA	GCTCACTGTA	GCCTCCTAGA	TTCAAGTGAT	CCTCCTGTCT	CAGACTCCTG	600
AGTAGCTAGG	ACTACAGGAA	CATTCCACCA	TGCCCAGCTA	ATTTTGTTTT	GTTTTGTTTT	660
GTTTTTCAGAG	ACAATGTATT	GCAGCGTTGC	CCAGGCTGAT	CTGAAACTCT	TAGCCTCAAA	720
CGATACTCCT	GCCTCAGCCT	CCCAAAGCAC	TAGGATTACA	GACATGAGCC	AATGCGCCCC	780
GCCTTAAATT	AGACTTTAAA	TGTGGTTTTA	AACTCCTGTT	GAAAAAGCGT	CTGGTATCTT	840
GAACCAGTAG	ATGTTTTTCAT	AGCAATGAAG	CTAAACTGTA	ATTTAGACAG	TAGCCAAATG	900
CTTGTGAAAT	TTTGCTAAAT	AATATAATCT	TCAAGGGAGC	AAATCATGTC	CCAAATGCAA	960
AAGATCAACT	GGTGGGGGCA	GTAGTAAAG	ACAGGATACT	GTGCTCTTTA	AAAGGTCAGT	1020
AACATAGTA	CCTAGTTATC	TTACTTATCA	CAGCAAAATA	ATTACATAAA	ATCCTATGGA	1080
TCATAAAGGC	ACAGACTCAC	TTCTGTCTCT	AGATCTCAAG	CTACCAAAAA	GAAATCTCCC	1140
AATAGTTTCT	TGGAGGCCTA	TACTTAGTGA	AAAAGCAGCT	GGAATCAACA	TAGTTCCTCC	1200
TATGTTGTAG	GACAATCCTA	GCTCTGGGCA	TACGAATACA	TTAAATCCCA	CTTATCTATA	1260

GAGCTTTCTT	AAAGGGAAGA	AATTTGAGTA	GATGTAAAA	CAGAATAAAA	GATTAAGGCT	1320
CCATAGGCAT	ACAGCTTACC	TCCAATTCTC	TTGGCCTCTT	GCAATTTCTA	TTATCAGGCT	1380
TTACAAGGTG	ATTTGCCATC	ATATTCCGAA	GGCACCAGCT	ACAAAGCTTA	GAACAATGCC	1440
AGATTTAGGT	ACAAACTCCA	TGCTACAAGC	TCTCTGGAAT	CCTTCCCTGT	TTCCCACTCC	1500
TACTGCTGAT	GTTAATTTAG	ACTGTCATTA	TCTGTCACTT	TCCTAAACTC	AATTTCTCCC	1560
TCCTCTAAAT	CATTCTATCA	ACTGCTATTT	GGGTAATCTT	TCAAAACCTT	GATTACTGCA	1620
TTCTTTTAAC	TCAAAAACCTT	TCATTGTTCC	AGAATAAGTT	GAAATTCCAT	GATATGGCCT	1680
TCAAGGTCCT	GTATTATCTG	GTGCAAGCCT	ACTAGTCCCA	TCATTTTCAA	CTACTCCTCT	1740
CTATGTACTT	AGCCAAATGA	GTCTCTCTGG	CAATTCTGCC	TTGTTTCAGG	ACTGGCTCAG	1800
TTAAGATTCT	TTTATCTTCG	GCCGGGCGCG	CTGGCTCACG	GCTGTAATCC	CAGCACTTGT	1860
GGAAGCTGAG	GCAGGAAGAT	CACCTGAGGT	CGGGAGTTCG	AGACCAGCCT	GGCCAGCATG	1920
GTGAAACCCT	GTGTCTACTA	AAAATCCAAA	CATTAGCCAG	GCGTGGTGGC	AGGCGCCTGT	1980
AATCCCAGCT	ACTTGGAAG	CTGAGGTGAG	AGAATCGCTT	GAACCCAGGA	GAGGGAGGTT	2040
GCAGTGAGCC	GAGATTGTGC	CATTGCACTC	CAGCCTGGGC	AACAGAGCGA	GACTCCACCT	2100
CAAAAAAAA	AAGGATTCTT	CTATCTTCAC	AAAATCTTAA	TGTTTAAACA	GGTCTTACAG	2160
TTCATCTAAT	TCAATCTCAT	TTTTTACAAG	TGAGAAAACA	GGGACAGTGA	CGGTGGATCA	2220
AGTGACACCA	GTAAGACTGA	GCTAAATTAG	AACCGAGATC	TCACTCGAGT	CTGAGGTTAT	2280
TCCCCTGTC	CAACCTTACT	TTAAAGTAGC	TTCAAATTTT	ACTTTTACTT	TTCCATAAAT	2340
TCGGAAGGGA	TTTTCCCTAG	GAGTCCAAAT	GTTGAAACCT	GGAAGGGTAT	AGTCTCTGTG	2400
TCTTTGAGAT	GAGGGGAGCC	CTGTCCATAT	TCAAGTTATC	AATTGACTTT	GTTGTTTTTG	2460
AGAAACGATG	CTGATTTGGG	TAACTTTAAAC	ACATCTGTTT	GATTAGTCCT	ATAAAATATG	2520
CATATATAGA	AGACAGAAAG	AGCAACAACA	AATTTGAAAG	ATGCTTGTTA	AGTAAATTCT	2580
GTATCGTACG	TGTCCATTCC	TGCCAGTACC	TTTATAGTAT	GTAAGTTTAC	GTGCTGTAAT	2640
AGTATTAATA	GTATCTAGAA	AATACTACAC	ATGCACAGGA	GTGCTAACTT	TGCCTTGGGA	2700
GTTGGAGAAAT	ACTTCAGAGA	AGCCAACAGG	CAGATTTTTC	TCTCTTCCCT	TCCCCTTCTA	2760
ATTTTCCCTT	TCCCCCTCAC	CCCCCTCTCT	TCTCTCCCCA	AGTAACACTG	TGCACCTATG	2820
TCAAACGAAA	ACTTATAATC	AAGTAACTGT	TTCTGCAAAA	ATAAGTTCGT	TTTCTGTGCA	2880
TGGCTCAAGG	CCTCAGCAGA	TCCAGGCCTG	GTGGACGGGC	TGGTCTTCGT	CGTGTGCCAA	2940
ACACTGACCA	CTGCCCTGGC	TCTGCCATCT	TAGGCTTAGT	GACCTGGCTG	TTACTAAGCA	3000
CTGTCCCCTC	TGCCCCATGC	AGCTGTCTCC	TTCTAGTCTT	CTCCCCTCTC	TCAACGCGAT	3060
CCTAGCCCCCT	CAGGCATTTT	CACCTCCATT	TTCCCTCACT	TCCCGCCGCC	TCCCGCACT	3120
TCCTCCCTAC	TGTTGTTTCC	GCCCCACTAG	AGCCCCCTCAG	AGAAAGTTTC	CATCCTCGCA	3180
CCCTTCCTTG	TGTCACAGCC	CGTCACATTC	TCACAGGCGC	CCATCCCTCC	AGCCCCACCC	3240
CAAGGCCAAT	GTACTTCGCG	GTATGGGGAC	CTTCTCTCGT	AGCGAACGCG	AGGGAGTGAA	3300
GACCTGGGGC	GCGGGGTGCT	CGGACTTCGG	GGGTGGAGGT	GGAAGCGCG	CCGCACTCCC	3360
AGCAGCCCCCT	GCACGAGTCA	CGTGACAGCT	CTCCACCAC	CACCCCCCCC	AACTTCCCCA	3420
CCGTAGCCTC	CCAGAGCCAG	GCCCCACGGA	AAGGCAGCTT	TTTCCCGGTT	TTCTCCCGCT	3480
CTTTCCCTC	CACTTGGAAT	ACTCGTGAAG	CAAAAATCTC	TCCCTGCCAC	CCTGTGTGTG	3540
TTTGAACCAG	GAAAAAATCT	GAAACTGGTC	AAGAAAGAAC	AAGGAAGACT	TGCCAAAGCA	3600
AGGCCGGTGT	GTGTCCACAG	AGCTTAGAAT	CTCAGCAAAAG	GAACACAAAA	TAGCACATCC	3660
ACGGCCTCTT	TTTCAGTAAA	ATTTACTTGG	TTTGTTTGCA	GGAAGGGTTT	AAAAGTGCCT	3720
TTGCAGATGC	TCTGTTTGCA	GGAAGGCTTT	AATCACGTGT	TCCCCTGGCC	CACAAGCAAG	3780
GCTTTTGTAG	CCAGAGCCTC	AGTTACTGCC	CCCTCTTCCT	CTTTGGTGCA	ACCAAACGTT	3840
CAGAATACAG	CCTTCTTAGA	AAATTCCTAC	CCCGGGTGTG	TCAATAAGTT	AAGTCTAATT	3900
GGCAACAGCT	ATCAAAAAGT	GTTGCATAAC	ACACATGGCT	CACATAATTG	TAGCTTTGCC	3960
TCATCGGGTG	TTTTAATGCG	GAGGCTTTGA	CCTGCAATTT	CAAAGATATA	CATTCCAAGC	4020
TTACGCCCAG	TTAGTGATG	TGGAAGAAAA	AAAAAAGCAA	ATTACCTCAT	AACACAAAGG	4080
TCAATAACAC	ACATCCATAA	GCTCCAGGTA	CAAAATCTTA	CATCTTAGAG	AACTATATTT	4140
AACATTTTAA	TACATTACTA	AGGTTTTTTT	TTTCCTTTTG	CTTGATTAAA	TGTTAGTTAT	4200
CATTAAAGTCT	TGGAATTATT	CTGTGTGTGT	ATTTTTATTT	GCTGTTTGTG	AAGAAGCCGG	4260
TTGTTTTTAA	TAAGTTCCTA	GAAAATAAGC	GCTCAATGTG	TTTAATCTGA	GTTGCTAATA	4320
TTGTGAAATA	TAGGCCACAT	AATACTAGCC	TAGATAACTA	TGGCGAAGTA	AGGAGTCTCA	4380
AACACTGTCC	CAGAACAATA	GCAATCTGTG	TTGAATTTT	ACCCTCTGTG	GTAAAATGAA	4440
GGGAAAAGGA	ATGAAGTTTT	AGTTTGCCTT	AATTTTTATC	TTTATTGTTT	CAGACTCTTC	4500
AGCAGTATAA	AGTTTTTCATC	AAGTCAAATA	TATTCACCTT	AAAGTGACTG	TGCTTTATTC	4560
TGATACCATG	TCCTTCCTAA	TTTGGGGGGC	CAGGTGAGAT	AAGTTTTATG	AAATAAAAAAG	4620
ATTAAAAAAT	CTTACATTTT	TAGTGTCTCT	CCTTGGTAAA	ATGTAGAGTT	GTCCACTGTG	4680
TTTATCTCCT	CCTCCTTATT	ATCATGGTTG	CTGTTATTAT	TTTTAATGGT	TCATTAAACC	4740
CAAGGGTCTG	GGAAATACTC	ATGGAATTCA	TCTCACAGCC	TTACACTGTG	ATGATATTTA	4800
AACAGGTGGT	TGTCCATCTG	ATTCTTAAAG	TATTTCCAAAG	AAAAATGATT	CCACCTAATG	4860
CATAAATGCT	TTCATCAGAT	TAAGAGAACA	CCATGGACAT	TTTATTTTAT	TTTATTTTTT	4920
AAATATTAAC	TTCCATTGCA	TAAGCTAAAT	GGGTAGGAAT	AAGTGAGATG	ATATTGTTAT	4980
CTAGAGCTTT	AAAATATTCA	AAGGGCTGTC	ATCATTATCT	CATTTAATCT	TTGAAAACAA	5040
CTCTATGAAG	TACAAAGGAC	ACTGAGACAT	TGTTGCTCT	ATATCAAAGA	AAAAAGTGT	5100
TGTCCCAAAA	CTTCAAAATG	TGTAAATTAC	ACATTTCTGCA	TCTTTACAGC	TGGAGAAAAT	5160
TCACTGGCAA	TGGAATATTT	AAAATTAGAG	CTTGCTTAGT	GTGCTGCTTC	TGATCACTAC	5220
TTGATCCAC	TTCGTGCTTT	CATGTTAATT	GGCCCAATTG	GACTCTACAG	TTGGAAGGTG	5280

AAAACCTTACT	ATTTCAACTT	GAGTCACGTA	TGTATTCTTA	TCATATACTT	CTTAAAGGTA	5340
CTATTTTTTTT	TCTTCTGATA	GTCACCACAC	CAAGCACTTC	CAGCCACCCT	GCCACAGACT	5400
TCCTTTGTAA	TCACTGTTGA	AGGACATGAT	GTTTTTATGA	CTTCCCAGAA	TGAAAACCTT	5460
ATCTTGTTTT	TAAAACAAAC	AAACCAACAA	AAAGTAGTGT	TTATGTAAGC	ATTTGTGTTCC	5520
CTGACTCTAG	GAACCCCTCT	GTTTTTATAT	CAACTCTGTA	CTGGCAAAAC	ACAAAACAA	5580
AATGCCACCT	TGCTAATTCC	CTTCTAGCA	AAGTAATACA	GTTTAGCACA	TGTTCAAGAA	5640
AAAAATGGCT	AAGAAATTTT	GTTTCCACTA	ATTATTTTCA	AGACTGTGAT	ATTACACTC	5700
TGCTCTTCAA	ACGTTACATT	TTATAAGACT	ATTTTTTAAC	ATGTTGAACA	TAAGCCCTAA	5760
ATATATTGGA	CCTTAAATTG	TATTTCAAAT	ATTTTAGGTC	AGTCTTTGCT	ATCATTTCCAG	5820
GAATAGAAAG	TTTTAACT	GGAAACTGCA	AGTAAATATT	TGCCCTCTTA	CCTGAATTTT	5880
GGTAGCCCTC	TCCCCAAGCT	TACTTTCTGT	TGCAGAAAGT	GTAAAAATTA	TTACATAAAA	5940
TTCTAATGAT	GGTATCCGTG	TGGCTTGCAT	CTGATACAGC	AGATAAAGAA	GTTTTATGAA	6000
AATGGACTCC	TGTTCCACTG	AAAAGTAAAT	CTTAATGGCC	TGTATCAACT	ATCCTTTGAC	6060
ACCATATTGA	GCTTGGAGG	AAGGGGAAGT	CCTGAATGAG	GTTATAAAGT	AAAAGAAAAT	6120
ATTTGCAAAA	TGTTCTTTT	TTTAAATGT	TACATTTTAG	AAATATTTTA	AGTGTGTAA	6180
CATTGTAGGA	ATTACCCCAA	TAGGACTGAT	TATTCCGCAT	TGTAAAATAA	GAAAAAGTTT	6240
TGTGCTGAAG	TGTGACCAGG	AAGTCTGAAA	ATGAAGAGAG	ACAGATGACA	AAAGAAGATG	6300
CTTCTAATGG	ACTAAGGAGG	TGCTTTCTTA	AAGTCAGAAA	GAGATACTCA	GAAAGAGGTA	6360
CAGGTTTTGG	AAGGCACAGA	GCCCCAECTT	TTACGGAAGA	AAAGATTTC	TGAAAATAGT	6420
GATATTACAT	TAAAAGAAGT	ACTCGTATCC	TCTGCCACTT	TATTTGCGACT	TCCATTGCC	6480
TAGGAAAGAG	CCTGTTTGAA	GGCGGGCCCA	AGGAGTGCCG	ACAGCAGTCT	CCTCCCTCCA	6540
CCTTCTTCTC	CATTCTCTCC	CCAGCTTGCT	GAGCCCTTTG	CTCCCTGGC	GACTGCCTGG	6600
ACAGCTCAGCA	AGGAATTGTC	TCCCAGTGCA	TTTTGCCCTC	CTGGCTGCCA	ACTCTGGCTG	6660
CTAAAGCGGC	TGCCACTTGC	TGCAGTCTAC	ACAGCTTCGG	GAAGAGGAAA	GGAACCTCAG	6720
ACCTTCCAGA	TGCTTCTCTC	TCGCAACAAA	CTATTTGTCTG	CAGGTAAGAA	ATATCATTCC	6780
TCTTTATTTG	GAAAGTCAGC	CATGGCAATT	AGAGGTAAAT	AAGCTAGAAA	GCAATTGAGA	6840
GGAATATAAA	CCATCTAGCA	TCACTACGAT	GAGCAGTCAG	TATCAACATA	AGAAATATAA	6900
GCAAAGTCAG	AGTAGAATT	TTTTCTTTTA	TCAGATATGG	GAGAGTATCA	CTTTAGAGGA	6960
GAGGTTCTCA	AACTTTTTCG	TCTCATGTTC	CCTTTACACT	AAGCACATCA	CATGTTAGCA	7020
TAAGTAACAT	TTTTAATTAA	AAATAACTAT	GTACTTTTTT	AACAACAAAA	AAAAGCATAA	7080
AGAGTGACAC	TTTTTTATTT	TTACAAGTGT	TTTAACTGGT	TTAATAGAAG	CCATATAGAT	7140
CTGCTGGATT	CTCATCTGCT	TTGCATTGAT	ACTACTGCAA	TATTGCACAG	AATGCAGCCT	7200
CTGGTAAACT	CTGTTGTACA	CTCATGAGAG	AATGGGTGAA	AAAGACAAAT	TACGTCTTAG	7260
AAATTATTAGA	AATAGCTTTC	ACTTTAGGAA	CTCCCTGAGA	ATTGCTGCTT	TTTGCTGGTA	7320
AGATAAATAA	GCTTCTCTTT	AAACGGAATC	TCAAGACAGA	ATCAGTTACA	TTAAAAGCAA	7380
ACAAAAAATT	TGCCCATGGT	TAGTCATCTT	GTGAAATCTG	CCACACCTTT	GGACTGGGCT	7440
ACAATTGGAT	AATATAGCAT	TCCCCGAGAT	AATTTTCTCT	CACAATTAAG	GAAAGGGCTG	7500
AATAAATATC	TCTGTTTGAA	GTTGAATAAC	AAAAATTAGG	ACCCCTAAA	TTTTAGGGCT	7560
CCTGAAGATC	GCTTTTTTGC	CTATATTCAG	CTACTTTACG	TTCTATTAAA	TTTCTTTTCA	7620
GGCCAGGTGC	ACTAGCTCAT	GCCTAGAATC	TCAGGCAGGC	CTGAGCCCAG	GAATTTGAGA	7680
CCAGCCAGGG	CAACACAGTC	TCTACAAAAA	AATAAAAAAT	TACCTGGGTG	TGTTGGTGCA	7740
TGCCTGTAGA	ACTACTCAGG	ATGCTGAGGA	CTGCTTGAGC	CCAGGATAGC	CAAATCTGTG	7800
GTGAGTTTCA	CCACTAAACA	GAGCGAGACT	TTCTCAAAAA	AAACAAACAA	AAAACCTAGC	7860
AACTTCTTTC	AAAATAACTT	TTTATCTGCA	ATGTTTTCTT	ATTGCCTGTG	AGATTAAAT	7920
TACTCTTTTA	CCTGATTTCC	AAAGCCCTCC	ATAATCTAAT	CCGACTTTAC	CTTGTGTTCA	7980
CTGCAAAATA	GCAGGACTGT	TCCACTACAA	TCCAAAAATC	ACAGGTTGGG	TGCAGTGGCT	8040
CACTCCTGTA	ATCCCAACAC	TTTGGAAGGC	CAAGGCAGGT	GGATTGCTTC	AGCTCAGGAG	8100
TTCAAGACCA	GCCTGGGCAA	CATGGCAAAA	ACCCTGTCTC	TCCAAAACAT	ACAAAATTA	8160
GCCAGATGTG	GTAGTATGTG	CCTGTATGCC	CAACTACTCA	AAAGGCTAAG	GCAAGAGGAT	8220
CACTTGAGCC	CAGGAGGTCA	AGGCTACAGT	GAGCCATGTT	TACTGTGTCA	CTGCACTCCA	8280
GCCTGGGTGA	TAGAGCAAGA	CCATGTCTCA	AAAAAAAAAA	AAAGAAAAGA	AAAGAAAAAA	8340
ACATCGCTCT	ATTAGTTTCA	CCCCCACCAC	AACATTGTTT	TGATTATCAC	ATAAATGCTG	8400
GTCCATTGCC	TTCTCTATCT	ATTCAAATCT	TTAAGCATTC	TTTGAGATT	AACTCAATTC	8460
TCCTTTTCAA	ACTAGGCCAT	TTAAACTACA	TCAGTTCCAT	TTTGATTTTC	TTGCTTTGAG	8520
TCTACAGACT	CAAAAACAAA	AACTTAAAAA	CTTATTTTTC	AAGTTTCTG	CTACTCTCAC	8580
TTCTTCAACA	CTCACATACA	CGCATTCTCA	ATAAGATGGC	AGAATGTTCA	AGGATAAAAT	8640
GATTTATAGA	ACTGAAAAGT	TAGGTTTTGA	TCTTGTGTGT	GTCAAGATGA	CTACCTACCT	8700
GATCTCAGGT	AATTAATTAT	GATGATGCT	CCCTCATTTT	ATCCCATACC	TATTCAACAG	8760
GATTGGAAAT	CCACAGCAAG	GATAAACATA	ATCATAGTTG	CTTTTCAAGT	TCAAGGCATT	8820
TTAACTTTTA	ATCTAGTAGT	ATGTTTGTG	TTGTTGTGTT	TGTTTGAGAT	GGAGCCCTGC	8880
TGTGTCACCC	AGGCTGGAGT	GCAGTGGCAC	GAACTCGGCT	CACTGCAACC	TCTGCCTCAT	8940
GGGTTCAATC	AGTTATTCTG	CCTCAGTGTC	CCAAGTAGCT	GGGACTACAA	GGCAGATGCC	9000
ACCATGCCCT	GCTAATTTTT	GTATTTTTC	TAGAAACAGG	GCTTCACCAT	GTTGGCCAGG	9060
CTGGTCTCGA	ACTCCTGACC	TCAAGTGATC	CAGCCGCTC	GGCCTCCCAA	AGTGCTGGGA	9120
TTACAGGCAT	AAGCCACCGT	GCCCAGCCTA	ATAGTATGTT	TTTAACTCT	TAGTGGCTTA	9180
ACAATGCTGG	TTGTATAATA	AATATGCCAT	AAATATTTAC	TGTCTTAGAA	TTATGAAGAA	9240
GTGGTTACTA	GGCCGTTTGC	CACATATCAA	TGGTTCTCTC	CTTACAGCTT	TAATTAGAGT	9300

CTAGAATTGC	AGGTTGGTAG	AGCTGGAACA	GACCTTAAAG	ATTGACTAGC	CAACTTCCTT	9360
GTCCAAATGA	GGGAAGTGA	ACCCTTAAAA	TTAAGTGACT	TGCCCCAGAC	AAAACTGGAA	9420
CTCATGTGTC	CTAATTTCCA	TCATGAAATT	CTACCATTGA	CTAGCCTCTG	GCTAGTTGTC	9480
AAAGTATTGC	ATAACTAAAT	TTTTATGTCT	TTTTTAAAGA	ACAAATTGTC	ACTGCTTACT	9540
CCTGGGAGGG	TCTTTCTGAG	GTGGTTTATA	ACTCTTAAAA	AAAAAAAAGT	CAGTAGTCTG	9600
AGAATTTTAG	ACGAAATAGT	CAAAGCATT	TTATCCAATG	GATCTATAAT	TTTCATAGAT	9660
TAGAGTTAAA	TCAAAGAAAC	ACGGATGAGA	AAGGAAGAGG	AAAATTGAGG	AGAGGAGGAA	9720
TGGGGATGAG	AACACACTAC	TTGTAATCAG	TCATAGATGT	ACTGAGAACT	AACAAGAAGA	9780
ATTGTAAGAA	AATAAGAATG	AAGAATTCAA	AATCAACACA	TGAAATAAAA	AGAAACTACT	9840
AGGAAAAAAT	GGAGAAAGACA	TTAGAAAAAT	TATTCTATTT	TTAAAAATTCT	GTTTTTCAGGC	9900
TTCCCTCCTG	TTCTTCCTCC	TTCTCATTGG	TTTTTCAGGTG	GAGGGAAAAGT	TTAAGATGGA	9960
AAAAATATAT	ATATTCTACA	CATCCCTTTC	TACGCTGTTG	TCATGGCAAC	AAGGTTTATC	10020
ATAGCAAAC	TTTATTTCATA	CAACATTTAT	TGAGTTCTTA	CTGTGTGGTA	AGCTCTTTCC	10080
AGGTGTTGAA	AATTCAGGGG	AAAAAAGACA	ACTCATTTGC	TTAAAACTCA	GATGAAAGCT	10140
GAACAGACCT	ATTTTAAATC	AAAGTAATCT	CAATTGAGGG	TAGTAAGAGC	TATTTAAGAA	10200
GCATGAACAG	GTGTGAAGGA	GGTAGGACTC	TGAGGAGAGA	ATAGTTAGCT	AGGAATGAAA	10260
GAGCAGAGAA	GTTTTCTCTAG	AGGAACATTT	AAAGCTGGGA	GTTACGGGAT	GAAAGATGAG	10320
GCAGGGTTTG	CAGGCAAAAA	AAAAAAAAG	GCAGGGGAAG	GGGAAGTTCT	GGCCTGGCAG	10380
AGAGATAAAC	TGTGGCAACA	ATGGAGGAGA	GTCTGGAAGC	AAGAAAACCA	AGTAGAAGAG	10440
TATTAAAAATA	GAAGATGCCA	GGGGTAATGA	GGGCTTGATT	TAAAACAGTG	CTGTTGGAGA	10500
TGGAGAGGAG	ATACCAAATT	CTGGAGACAT	TTCTGAGTTA	GAACCTACAG	TATTTATCAG	10560
ACAAGGGAAA	GATTAGACAA	AGGAGTTAAG	AATGACTCCC	AGGTTTCAGT	TTGGGGCAGG	10620
TAAGTAGGAC	ATGTTTTGAA	AAGTAATGTA	TTGGATCTCT	TACCATTGGA	ACTATGTATG	10680
TGGAGCCAAA	TTAAAATTTG	TACATGTATA	TAATCTCCC	CCCACCACCA	GTAACCTACT	10740
CCCTAACTCT	CTACTTTGTA	GCCAGACTTC	CTAAAAGAAT	AGTTTGTAGT	CACTGTCTTT	10800
ACTTTTCCCC	TCCCATTCTG	TCCTAGATAT	TTGTCCACCT	ACCATCTGCT	GCCTCCACTT	10860
TACCCAAACT	GTTCTACGGT	TGCCCCAAAC	TTCCTAATTG	CCAAATTCAA	TGAACAAGTT	10920
TAAGCTTATA	TGTAATTTAG	GAGCTCTACA	GTTTGATTTT	GAGCAGCCCC	TCCTGAAACC	10980
CTTTCTCTTT	CGACTTCTGT	GACACATCTC	AGATTTCACA	AACTGAACTA	ATTATTTTAC	11040
ACTTGAGCTG	TATTTTCGTT	CTTCTTTCTT	GATGAATGAG	GTAACCACTC	AACAAATTGC	11100
CCAAGCCAAA	AACTACGAAG	TCATCCTCAG	TTCCTCCTTC	TTCTGTTTGA	CCCACAACAG	11160
ATCAGCTGAG	AAATCCCCTG	GTTTAGTATC	TCTTGAATTC	ATTACCTTAA	TTTATAGCCT	11220
CATCAACTCT	TAATTGTTAA	AATTACTTCA	GTAGTTGTTG	TCTGACCTCT	GTCCAATCTT	11280
GTTCAATCAG	TGCTTCTCTT	TTGTTCTTGG	TGGTGGTGGT	GGTGTGACAA	GGTTCCTGCT	11340
TTTGCTGCCC	AGGCTGAAGT	GCAGTGGAGC	ACTTCACTGC	AACCACAGCC	TCCTGGGTTT	11400
AAGCAGTTCA	CCCTCCCGAG	TAGCTGGGAC	TACAGGTATG	TGCCACCACA	CCCAGCTAAT	11460
TTTGTGTTTT	CAGTAGAGAC	AGGGTTTCAC	CATGTTGGTC	AGGCTGGTCT	CAAACCTCCTG	11520
ACCTCAAGCA	ATCCACCCAC	CTCAGCCTCC	CAAAGTGCTG	GGATTACAGG	CATGAGCCAC	11580
TGCACACGGA	CCAGATCCAT	TGTTTATGTT	GCTTCTAGAG	TGAGTTTTTA	AAACACAAAT	11640
TTGACCATAT	CTTTCTCCAA	TTTAAGTCAG	TATTTTTTTT	TTCAGGAAAA	AACAGTTCAA	11700
ACTCTTTAGT	CTGCTTACAC	AAGGCCTTTG	TAGTCTGACT	CTTCTTTCCA	AGCTTTCATC	11760
AAAGTATACT	GCAAGTTACA	TTTATGTGTA	ATTGAATTAG	GCAACGGTAT	AAAAATTATA	11820
GTTTATATGG	GCAAAATGGA	AATAATGTTA	ACTCTTCCAA	ATAGTTTATC	TAGAATTGCA	11880
TAATTTTCAA	GCTGTGAGGT	CAAATGAGTT	ATAAATAGGC	AACACTATTG	CCACATGCAA	11940
GTGTCTCTTA	TACTTGGTAG	AATTATCTGC	TTCCATGTCA	TTATTATGTA	AATTAGACTT	12000
TAAATAACTC	AGAAGTTCTT	CAGACATACA	GGTTATTATT	GTGCTTTTTA	AACATAATTT	12060
TAAATAATTT	TATATATGAT	AATGTTATCC	AAGTGCTAAG	GGATGTATTG	TTACTGCTGT	12120
GCAAAAAAAA	AAAAAAAAG	AACTCCAAAT	AAATATGTTG	AAACCAAGTT	TATATGCAAG	12180
AAAACAATAT	TAAAAAGGCC	AAAGTACCAC	CATAATAGGC	TGTGTGGAGA	CGGCAGGCTA	12240
CAAAACACTA	GTAATAATGC	TGAGAAAGTT	GAAAAAAGAA	AGAAAGCAAC	AATATGCTTT	12300
GGTTGTTGTA	GGTTTATGTA	CTCCAAGAAT	ATCTCCTCTC	AACTTTTAC	GTTTTTTCCA	12360
AAGAAAAGTT	AACTTTGGCT	GGGCGCAGTG	GCTCTTGCC	GTAGTCCCAG	CCTTTGGGAG	12420
GCCAAAGCGG	GCAGATCACC	TGAGGTCAGG	AGTTTGAGAC	CAGCCTGACC	AAAAATGGAG	12480
AAACCCGCCC	CCCTCACTAC	TAAAAAGAATA	CAAAATTAGG	CCGGGCACAG	TGGCTTACCC	12540
CTGTGATCCC	AGCACTTTGG	GAGGCCGAAG	CAGGAAGATC	ACCTGAGGTC	AGGAGTTCTG	12600
GACCAGCCAT	GGAGAAACCC	GTCTCTACTA	AAAATACAAA	ATTAGCCGGG	CGTGGTGGTG	12660
CATGACTGTA	ATCCCAGCTA	CTCAGGAGGC	TAAGGCAGAG	AATCACTTGA	ACCCAGGCAG	12720
TGGAGGGTTG	AGTGAGCCGA	GATCGTGCCA	TTGCACTCCA	GCCTGGGCAA	CAAGAGCCGA	12780
ACTCTGTATC	CAGAAAAACAA	AAGAAAAAGAA	AAGGTAACTT	TGAACCTATG	GAGATCTTTA	12840
GAAATGCATT	CTTTCTGTAA	AATGTGACTA	CATTTGCCTT	ATTTATGGTA	AAAATGTTGA	12900
GGCCTCAAAC	AACCCATATT	TTCTCGGTCT	CCCCGCTGCC	TAGCCTTTGT	TCACATTGCT	12960
TCTTCTTGGT	GGAAGCTCTT	CCTCTGGCCT	TGAAAATGCC	TGCTTCTCTT	TCAAGGTAGC	13020
ACAGTCAATC	CTTTCTGTGG	TAACCTTCTC	CAGCACCATC	AAACAGAAAG	AATGAATCTC	13080
TTGTAAATTC	AGCTCTTACG	TCATTCTTAC	CATTATTTTG	TAACCTTTTA	TAGATCTTTC	13140
TCTCCCACTA	GACTCTGAGT	CACTGGAGAG	TAGGAGCCAA	CTCTCATTCA	TGTGTGGTTT	13200
GGTCAGCTAC	TGGCCACATT	CCTGATGCAT	AGTTAATGCT	CAAACCTTAA	CTGGTGAATC	13260
AGCTCAAATA	TTGTCCTTCT	CTAAATCCAT	TCACTCATTG	ACTAACTATG	TACTCAAAAT	13320

AGTAAACACC	AGTAATTTAA	TCCAATTCCT	GCCCATACTG	CTTGGTACAT	TTCAGGTGAA	13380
TTAGTTTGAT	AAATATGTGT	GTATTACATA	ATATTAAAGT	ATGTACAGAA	GATCATGCTA	13440
ATCATAATTC	ACAACCTGATA	ACTAATCAAA	CATAAATGCT	CTCAGGTAA	CAAATGTCTG	13500
CCTTCTCAGT	TAATGCAGTC	ATTAACAAAC	ACCTTCTGAT	GCTGATAATA	GGGCCTTGTT	13560
CAGCAATGAA	GCCATAAAGG	TGAATAAAGA	ACATGCCCTC	GTGGAGCTCA	CAGCCTAGTC	13620
ATTATTGTTC	TGATTTTTAA	TATTAATGTT	GGTTTGGGTT	TTGGTGAAAA	ATGTTTAGAC	13680
TTATCTTAGT	GATCTTTTCA	TCCTTTGCTA	TATTATTTTT	CTCTAAGAGT	CTTCCTTATC	13740
CCCTCCTTTA	AAAAACTAGG	TGATAATTCT	AAATTGTAAA	TTTAAATATT	ATAAATAGCT	13800
TATAAAATTT	AATATTTATA	ATATTTAAAT	GTTTGATAAA	TATTTAAATT	TTATAATATT	13860
TAAATGTTTA	TTTAAATTC	TTTGACATC	AGTTTTTATT	TTATTTAAAT	GTGTTGGCCA	13920
GGCATGGTGG	CTGACACCTA	TAATCCCAGA	ACTTTGAGAG	GCCAAGTCAG	GCAAACCATT	13980
TGAGCTCAGG	AGTTTGAGAC	CACCCTGGGC	AACGTGGTGA	AACCCTGTCT	CTACCAACA	14040
TATGAAAACT	TATCTGGGTG	TGGTGGCAG	CATCTGTGGT	CCCAGATGGG	AGTCCCAGGC	14100
TAAAGATGGGA	GAATCGCTTG	AACCCAGGTG	AGAGGGGTGG	GGTGGATGTT	GCAGTTAGCT	14160
GAGATCGTGC	CACCTGCATC	CAACCTGGGT	GACAGAGTGA	GACTCCATCT	CAAAAAA	14220
AAATGTTATC	TAAATAAGAT	AAATTTAATA	ACTGTTTCGCA	CTTAGATGAG	CATAAGGAAC	14280
TAAACCTAGA	TAAACTATC	AAATAAGGCC	TGGGTACAGT	GACTCATGCC	TGTAATCTCA	14340
AGCACTTTGG	GAGGCCAAAA	TTATACAAAG	TTAGTTGTAT	AACACCAACT	AACAACCTATT	14400
TTGGGGTTAG	CTTAATTCAG	ATTAATTTTT	TTTAAACTGA	GTTTTAAATT	CCTGCTTACT	14460
CTACCATACA	TGCTAGGCCT	CATATTATGC	TAGAAAAATT	TTGAGCACAG	ATTTATTGAAT	14520
ACTCTCCTGC	ATACCATTTA	ATTTTTAAAC	AAATTTTAAT	GCAGTATATA	TGTGCCTTTT	14580
TACCAACACA	TTAAATAATA	AGATCTACTG	TGAGGACTAA	ATTTCTGTAA	TTTCAAAGTA	14640
GTAATGAGTT	TAAACCATGT	CTCAAGATCT	CTGCAATAAC	TGTAGCACAA	CAGAAAAATAG	14700
GTATTTCTAT	TAATGACAGA	GTCACAAGTA	CTACTAATAA	TACTGTGGTT	TGTTTCTGTC	14760
AACTAATCAT	GGGAGGAATG	CTAAATTTCA	GAGGTTGGTG	AAAATACATG	TGTATTTTTT	14820
TCCCCATCCA	AGTTCACAGA	TTTCTCACAC	TGAGAACTCC	TATTCCATAA	CAAAATTCTG	14880
GAAGCCTGCA	CACCGTATTG	GAAGAAGGGC	AGAAAGGAAA	AGCAAATGGA	AGGATTTAAA	14940
TTTTTTTCAA	ATCCTGTATC	CCTTGATTTT	ACAGCAAGAT	TGTATTTATG	TATTACTTGT	15000
GTTAAAAATA	TAGTATAATC	GAGACTCCAG	ATCAAAAATC	ACCGCAGCTC	AGCGAGAAAG	15060
AGGGCCACCA	AATGCCAGAG	CCCTCCAGCC	TTCTCCACC	CTGCCTGTAC	CCTCAGATGG	15120
AAGCACTTTT	TTATCATTGT	TTCACCTTTA	GCATTTTGAC	AATGAAGTCA	CAAACCTTCA	15180
GCCTCTCACC	CATAGGAACC	CACCTGGTGT	AAGAGAAGGA	TGAAGCCAGT	CCTTCCTAAA	15240
GGGCACGATT	AGATGTGTTT	ATGGCATCCT	CAGGTGAAAC	TATATTTATA	TTGACAATAT	15300
ATTTATATTT	CTCAAGGAAT	ACTAGAATAA	TGATTCAGTT	CAGTACTAGG	CACTTTATCT	15360
ACCCTTTATA	ATATTGTTTA	ATGAGAAAT	GCTTTCTATC	TTCCAAATAT	CTGATGATTT	15420
GTAAGAGAAC	ACTTAAACAT	GGGTATTCAT	AAGCTGAAAC	TTCTGGCATT	TATTGAATGT	15480
CAAGATTGTT	CATCAGTATA	CTAGGTGATT	AACCTGACCAC	TGAACCTGAA	GGTAGTATAA	15540
AGTAGTAGTA	AAAGGTACAA	TCATTGTCTC	TTAACAGATG	GCTCTTGCT	TTCATTAGGA	15600

ATAAAG ATG GCT GCT GAA CCA GTA GAA GAC AAT TGC ATC AAC TTT GTG GCA 15651
Met Ala Ala Glu Pro Val Glu Asp Asn Cys Ile Asn Phe Val Ala

-35 -30 -25
ATG AAA TTT ATT GAC AAT ACG CTT TAC TTT ATA G GTAAGGC TAATGCCATA 15702
Met Lys Phe Ile Asp Asn Thr Leu Tyr Phe Ile Ala

GAACAAATAC	CAGGTTTCTA	TAAATCTATT	CAATTAGAAA	AGATGTTGTG	AGGTGAACTA	15762
TTAAGTGACT	CTTTGTGTCA	CCAAATTTCA	CTGTAATATT	AATGGCTCTT	AAAAAATAG	15822
TGGAGCTCTA	GAAATTAACC	ACAACATGTC	CAAGGTCTCA	GCACCTTGTC	ACACCACGTG	15882
TCCTGGCACT	TTAATCAGCA	GTAGCTCACT	CTCCAGTTGG	CAGTAAGTGC	ACATCATGAA	15942
AATCCCAGTT	TTTATGGGAA	AATCCCAGTT	TTCATTGGAT	TTCCATGGGA	AAAATCCCAG	16002
TACAAAACCTG	GGTGCATTCA	GGAAATACAA	TTTCCCAAAG	CAAATTGGCA	AATTATGTAA	16062
GAGATTCTCT	AAATTTAGAG	TTCCGTGAAT	TACACCATT	TATGTAAATA	TGTTTGACAA	16122
GTAAAAATTG	ATTCTTTTTT	TTTTTTTCTG	TTGCCCAGGC	TGGAGTGCAG	TGGCACAATC	16182
TCTGCTCACT	GCAACCTCCA	CCTCCTGGGT	TCAAGCAATT	CTCCTGCCTC	AGCCTTCTGA	16242
GTAGCTGGGA	CTACAGGTGC	ATCCCGCCAT	GCCTGGCTAA	TTTTTGGGTA	TTTTTACTAG	16302
AGACAGGGTT	TTGGCATGTT	GTCCAGGCTG	GTCTTGGACT	CCTGATCTCA	GATGATCCTC	16362
CTGGCTCGGG	CTCCCAAAGT	GCTGGGATTA	CAGGCATGAA	CCACCACACA	TGGCCTAAAA	16422
ATTGATTCTT	ATGATTAATC	TCCTGTGAAC	AATTTGGCTT	CATTTGAAAG	TTTGCCTTCA	16482
TTTGAAACCT	TCATTTAAAA	GCCTGAGCAA	CAAGTGAGA	CCCCATCTCT	ACAAAAAATC	16542
GCAAAATATC	CTGTGGACAC	CTCCTACCTT	CTGTGGAGGC	TGAAGCAGGA	GGATCACTTG	16602
AGCCTAGGAA	TTTGAGCCTG	CAGTGAGCTA	TGATCCCACC	CCTACACTCC	AGCCTGCATG	16662
ACAGTAGACC	CTGACACACA	CACACAAAAA	AAAACCTTCA	TAAAAAATTA	TTAGTTGACT	16722
TTTCTTAGGT	GACTTTCCGT	TTAAGCAATA	AATTTAAAG	TAAAATCTCT	AATTTTAGAA	16782
AATTTATTTT	TAGTTACATA	TTGAAATTTT	TAAACCCTAG	GTTTAAGTTT	TATGCTAAAA	16842
TTACCTGAGA	ACACACTAAG	TCTGATAAGC	TTCAATTTAT	GGGCCTTTTG	GATGATTATA	16902
TAATATTCTG	ATGAAAGCCA	AGACAGACCC	TTAAACCATA	AAAATAGGAG	TTGAGAAAG	16962
AGGAGTAGCA	AAAGTAAAAG	CTAGAATGAG	ATTGAATTCT	GAGTCGAAAT	ACAAAATTTT	17022

ACATATTCTG TTTCTCTCTT TTTCCCCCTC TTAG CT GAA GAT GAT G GTAAAAGT 17075
Ala Glu Asp Asp Glu
-10

AGAAATGAAT TTATTTTTCT TTGCAAACTA AGTATCTGCT TGAGACACAT CTATCTCACC 17135
ATTGTCAGCT GAGGAAAAAA AAAAATGGTT CTCATGCTAC CAATCTGCCT TCAAAGAAAT 17195
GTGGACTCAG TAGCACAGCT TTGGAATGAA GATGATCATA AGAGATACAA AGAAGAACCT 17255
CTAGCAAAAG ATGCTTCTCT ATGCCTTAAA AAATTCTCCA GCTCTTAGAA TCTACAAAAT 17315
AGACTTTGCC TGTTCATTG GTCCTAAGAT TAGCATGAAG CCATGGATTG TGTGTAGGG 17375
GGAGCGTTGC ATAGGAAAAA GGGATTGAAG CATTAGAATT GTCCAAAATC AGTAACACCT 17435
CCTCTCAGAA ATGCTTTGGG AAGAAGCCTG GAAGGTTCCG GGTGGTGGT GGGGTGGGGC 17495
AGAAAATTCT GGAAGTAGAG GAGATAGGAA TGTCAAAGAA AAGAAGACCA CATTGAGAGG 17555
CCAAAAGCTG AAAGAAACCA TGGCATTAT GATGAATTCA GGGTAATTCA GAATGGAAGT 17615
AGAGTAGGAG TAGGAGACTG GTGAGAGGAG CTAGAGTGAT AAACAGGGTG TAGAGCAAGA 17675
CGTCTCTCA CCCCAGATG TGAATTTGG ACTTTATCTT GGAGATAATA GGGTTAATTA 17735
AGCACAATAT GTATTAGCTA GGGTAAAGT TAGTTTGTG TAACAAAGAC ATCCAAAGAT 17795
ACAGTAGCTG AATAAGATAG AGAATTTTTC TCTCAAAGAA AGTCTAAGTA GGCAGCTCAG 17855
AAGTAGTATG GCTGGAAGCA ACCTGATGAT ATTGGGACCC CCAACCTTCT TCAGTCTTGT 17915
ACCCATCATC CCCTAGTTGT TGATCTCACT CACATAGTTG AAAATCATCA TACTTCCTGG 17975
GTTCATATCC CAGTTATCAA GAAAGGGTCA AGAGAAGTCA GGCTCATTCC TTCAAAGAC 18035
TCTAATTGGA AGTTAAACAC ATCAATCCCC CTCATATTCC ATTGACTAGA ATTTAATCAC 18095
ATGGCCACAC CAAGTGCAAG GAAATCTGGA AATATAATC TTTATTCCAG TAGCCATAT 18155
GACTCTTTAA AATTCAGAAA TAATATATTT TTAATAATC ATTCTGGCTT TGGTATAAAG 18215
AATTGATGGT GTGGGGTGAG GAGGCCAAAA TTAAGGGTTG AGAGCCTATT ATTTTAGTTA 18275
TTACAAGAAA TGATGGTGTC ATGAATTAAG GTAGACATAG GGGAGTGCTG ATGAGGAGCT 18335
GTGAATGGAT TTTAGAAACA CTTGAGAGAA TCAATAGGAC ATGATTTAGG GTTGGATTG 18395
GAAAGGAGAA GAAAGTAGAA AAGATGATGC CTACATTTTT CACTTAGGCA ATTTGTACCA 19455
TTCAGTGAAG TAGGGAACAC AGGAGGAAGA GCAGGTTTTG GTGTATACAA AGAGGAGGAT 18515
GGATGACGCA TTTCTGTTTT GATCTGAGAT GTCTGTGGAA CGTCTAGTG GAGATGTCCA 18575
CAAACCTCTT TACATGTGGT TCTGAGTTCA GGACACAGAT TTGGGCTGGA GATAGAGATA 18635
TTGTAGGCTT ATACATAGAA ATGGCATTG AATCTATAGA GATAAAAAGA CACATCAGAG 18695
GAAATGTGTA AAGTGAGAGA GGAAAAGCCA AGTACTGTGC TGGGGGGAAT ACCTACATTT 18755
AAAGGATGCA GTAGAAAGAA GCTAATAAAC AACAGAGAGC AGACTAACCA AAAGGGGAGA 18815
AGAAAACCA AGAGAATTCC ACCGACTCCC AGGAGAGCAT TTCAAGATTG AGGGGATAGG 18875
TGTTGTGTTG AATTTTGCAG CCTTGAGAA CTAGTGAAAG CAGCTTGATG GTGAAATGGA GGCAGAGGCA 18935
CAACAAGGAG TTTGTTGATC TCACTGAAAG CAGCTTGATG GTGAAATGGA GGCAGAGGCA 18995
GATTGCAATG AGTGAAACAG TGAATGGGAA GTGAAGAAAT GATACAGATA ATTCTTGCTA 19055
AAAGCTTGGC TGTTAAAAGG AGGAGAGAAA CAAGACTAGC TGCAAAGTGA GATTGGGTTG 19115
ATGGAGCAGT TTTAAATCTC AAAATAAAGA GCTTTGTGCT TTTTGTGTTA TGAAAATAAT 19175
GTGTTAATTG TAATAATTG AGGCAATGAA AAAAGATAAT AATATGAAAG ATAAAAATAT 19235
AAAAACCACC CAGAAATAAT GATAGCTACC ATTTTGATAC AATATTCTTA CACTCCTTTC 19295
TATGTATATA TACAGACACA GAAATGCTTA TATTTTTATT AAAAGGGATT GTACTATACC 19355
TAAGCTGCTT TTTCTAGTTA GTGATATATA TGGACATCTC TCCATGGCAA CGAGTAATTG 19415
CAGTTATATT AAGTTCATGA TATTTACAAA TAAGGGCATA TCTTTGCCCT TTTTATTTAA 19475
TCAATTTCTA ATTGGTGAAT GTTTGTTTCC AGTTTGTGTT GTTTATTAAC AATGTTCCCA 19535
TAAGCATTCC TGTACACCAA TGTTCACACA TTTGTCTGAT TTTTCTTCA GGATAAAACC 19595
CAGGAGGTAG AATTGCTGGG TTGATAGAAG AGAAAGGATG ATTGCCAAAT TAAAGCTTCA 19655
GTAGAGGGTA CATGCCGAGC ACAAATGGGA TCAGCCCTAG ATACCAGAAA TGGCACTTTC 19715
TCATTTCCCC TTGGGACAAA AGGGAGAGAG GCAATAACTG TGCTGCCAGA GTTAAATTTG 19775
TACGTGGAGT AGCAGGAAAT CATTGCTGTA AAATGAAAAA AGAGATGATG TTGTAGAGGT 19835
CCTGAAGAGA GCAAAGAAAA TTTGAAATTG CGGCTATCAG CTATGGAAGA GAGTGCTGAA 19895
CTGGAACAAA AAAGAAGTAT TGACAATTGG TATGCTTGTA ATGGCACCGA TTTGAACGCT 19955
TGTGCCATTG TTCACCAGCA GCACTCAGCA GCCAAGTTTG GAGTTTTGTA GCAGAAAGAC 20015
AAATAAGTTA GGGATTTAAT ATCCTGGCCA AATGGTAGAC AAAATGAACT CTGAGATCCA 20075
GCTGCACAGG GAAGGAAGGG AAGACGGGAA GAGGTAGAT AGGAAATACA AGAGTCAGGA 20135
GACTGGAAGA TGTGTGATA TTTAAGAACA CATAGAGTTG GAGTAAAAGT GTAAGAAAAC 20195
TAGAAGGGTA AGAGACCGGT CAGAAAGTAG GCTATTTGAA GTTAACACTT CAGAGGCAGA 20255
GTAGTTCTGA ATGGTAACAA GAAATTGAGT GTGCCTTTGA GAGTAGGTTA AAAACAATA 20315
GGCAACTTTA TTGTAGCTAC TTCTGGAACA GAAGATTGTC ATTAATAGTT TTAGAAAAC 20375
AAAATATATA GCATACTTAT TTGTCAATTA ACAAAGAAAC TATGTATTTT TAAATGAGAT 20435
TTAATGTTTA TTGTAG AA AAC CTG GAA TCA GAT TAC TTT GGC AAG CTT GAA 20486
Glu Asn Leu Glu Ser Asp Tyr Phe Gly Lys Leu Glu

-5 1 5

TCT AAA TTA TCA GTC ATA AGA AAT TTG AAT GAC CAA GTT CTC TTC ATT 20534
Ser Lys Leu Ser Val Ile Arg Asn Leu Asn Asp Gln Val Leu Phe Ile

10 15 20

GAC CAA GGA AAT CGG CCT CTA TTT GAA GAT ATG ACT GAT TCT GAC TGT 20582
Asp Gln Gly Asn Arg Pro Leu Phe Glu Asp Met Thr Asp Ser Asp Cys

	25		30		35		
AGA G	GT	ATTTTTTTTA	ATTCGCAAAC	ATAGAAATGA	CTAGCTACTT	CTTCCCATT	20638
Arg Asp							
	40						
TGTTTTACTG	CTTACATTGT	TCCGTGCTAG	TCCCAATCCT	CAGATGAAAA	GTCACAGGAG		20698
TGACAATAAT	TTCACCTTACA	GGAAACTTTA	TAAGGCATCC	ACGTTTTTTA	GTTGGGGTAA		20758
AAAATTGGAT	ACAATAAGAC	ATTGCTAGGG	GTCATGCCTC	TCTGAGCCTG	CCTTTGAATC		20818
ACCAATCCCT	TTATTGTGAT	TGCATTAAC	GTTTAAAACC	TCTATAGTTG	GATGCTTAAT		20878
CCCTGCTTGT	TACAGCTGAA	AATGCTGATA	GTTTACCAGG	TGTGGTGGCA	TCTATCTGTA		20938
ATCCTAGCTA	CTTGGGAGGC	TCAAGCAGGA	GGATTGCTTG	AGGCCAGGAC	TTTGAGGCTG		20998
TAGTACACTG	TGATCGTACC	TGTGAATAGC	CACTGCACTC	CAGCCTGGGT	GATATACAGA		21058
CCTTGTCTCT	AAAATTAATA	AAAAAAAAAA	AAAAAACCTT	AGGAAAGGAA	ATTGATCAAG		21118
TCTACTGTGC	CTTCCAAAAC	ATGAATTCCA	AATATCAAAG	TTAGGCTGAG	TTGAAGCAGT		21178
GAATGTGCAT	TCTTTAAAAA	TACTGAATAC	TTACCTTAAC	ATATATTTTA	AATATTTTAT		21238
TTAGCATTTA	AAAGTTAAAA	ACAATCTTTT	AGAATTCATA	TCTTTAAAAA	ACTCAAAAAA		21298
GTTGCAGCGT	GTGTGTTGTA	ATACACATTA	AACTGTGGGG	TTGTTTGT	GTTTGAGATG		21358
CAGTTTCACT	CTGTCAACCA	GGCTGAAGTG	CAGTGCAGTG	CAGTGGTGTG	ATCTCGGCTC		21418
ACTACAACCT	CCACCTCCCA	CGTTCAGCG	ATTCTCATGC	CTCAGTCTCC	CGAGTAGGTG		21478
GGATTACAGG	CATGCACCAC	TTACACCCGG	CTAATTTTTG	TATTTTTAGT	AGAGCTGGGG		21538
TTTACCATTG	TTGGCCAGGC	TGGTCTCAAA	CCCCTAACCT	CAAGTGATCT	GCCTGCCTCA		21598
GCCTCCCAAA	CAAAACAAAC	ACCCACAGT	TTAATATGTG	TTACAACACA	CATGCTGCAA		21658
CTTTTATGAG	TATTTTAATG	ATATAGATTA	TAAAAGGTTG	TTTTTAACTT	TTAAATGCTG		21718
GGATTACAGG	CATGAGCCAC	TGTGCCAGGC	CTGAAGTGTG	TTTTTAAAAA	TGTCTGACCA		21778
GCTGTACATA	GTCTCCTGCA	GACTGGCCAA	GTCTCAAAGT	GGGAACAGGT	GTATTAAGGA		21838
CTATCCTTTG	GTTAAATTTT	CGCAAATGTT	CCTGTGCAAG	AATCTTCTTA	ACTAGAGTTT		21898
TCATTTATTA	TATTTATTTT	AG AT AAT	GCA CCC CGG	ACC ATA TTT	ATT ATA		21949
		Asp Asn Ala Pro Arg Thr Ile Phe Ile Ile					
	40			45			
AGT ATG TAT AAA GAT AGC CAG CCT AGA GGT ATG GCT GTA ACT ATC TCT							21997
Ser Met Tyr Lys Asp Ser Gln Pro Arg Gly Met Ala Val Thr Ile Ser							
50	55	60	65				
GTG AAG TGT GAG AAA ATT TCA ACT CTC TCC TGT GAG AAC AAA ATT ATT							22045
Val Lys Cys Glu Lys Ile Ser Thr Leu Ser Cys Glu Asn Lys Ile Ile							
	70	75	80				
TCC TTT AAG GTAAGACTG AGCCTTACTT TGTTTTCAAT CATGTTAATA TAATCAATAT							22103
Ser Phe Lys							
AATTAGAAAT ATAACATTAT TTCTAATGTT AATATAAGTA ATGTAATTAG AAAACTCAAA							22163
TATCCTCAGA CCAACCTTTT GTCTAGAACA GAAATAACAA GAAGCAGAGA ACCATTAAAG							22223
TGAATACTTA CTAAAAATTA TCAAACCTCT TACCTATTGT GATAATGATG GTTTTTCTGA							22283
GCCTGTACAC GGGGAAGAGG AGATACAACA CTTGTTTTAT GACCTGCATC TCCTGAACAA							22323
TCAGTCTTTA TACAATAAT AATGTAGAAT ACATATGTGA GTTATACATT TAAGAATAAC							22403
ATGTGACTTT CCAGAATGAG TTCTGCTATG AAGAATGAAG CTAATTATCC TTCTATATTT							22463
CTACACCTTT GTAAATTATG ATAATATTTT AATCCCTAGT TGTTTTGTG CTGATCCTTA							22523
GCCTAAGTCT TAGACACAAG CTTACAGCTT CAGTTGATGT ATGTTATTTT TAATGTAAAT							22583
CTAATTGAAT AAAAGTTATG AGATCAGCTG TAAAAGTAAT GCTATAATTA TCTTCAAGCC							22643
AGGTATAAAG TATTTCTGGC CTCTACTTTT TCTCTATTAT TCTCCATTAT TATTCTCTAT							22703
TATTTTCTC TATTTCTCTC ATTATTGTTA GATAAACCAC AATTAACATAT AGCTACAGAC							22763
TGAGCCAGTA AGAGTAGCCA GGGATGCTTA CAAATTGGCA ATGCTTCAGA GGAGAATTCC							22823
ATGTCATGAA GACTCTTTTT GAGTGGAGAT TTGCCAATAA ATATCCGCTT TCATGCCAC							22883
CCAGTCCCCA CTGAAAGACA GTTAGGATAT GACCTTAGTG AAGGTACCAA GGGGCAACTT							22943
GGTAGGGAGA AAAAAGCCAC TCTAAAATAT AATCCAAGTA AGAACAGTGC ATATGCAACA							23003
GATACAGCCC CCAGACAAAT CCCTCAGCTA TCTCCCTCCA ACCAGAGTGC CACCCCTTCA							23063
GGTGACAATT TGGAGTCCCC ATTCTAGACC TGACAGGCAG CTTAGTTATC AAAATAGCAT							23123
AAGAGGCCTG GGATGGAAGG GTAGGGTGGA AAGGTTAAG CATGCTGTTA CTGAACAACA							23183
TAATTAGAAG GGAAGGAGAT GGCCAAGCTC AAGCTATGTG GGATAGAGGA AAACCTCAGCT							23243
GCAGAGGCAG ATTCAGAAAC TGGGATAAGT CCGAACCTAC AGGTGGATTG TTGTTGAGGG							23303
AGACTGGTGA AAATGTTAAG AAGATGGAAG TAATGCTTGG CACTTAGTAG GAACTGGGCA							23363
AATCCATATT TGGGGGAGCC TGAAGTTTAT TCAATTTTGA TGGCCCTTTT AAATAAAAAG							23423
AATGTGGCTG GGCCTGGTGG CTCACACCTG TAATCCCAGC ACTTTGGGAG GCCGAGGGG							23483
CGCGATCACC TGAAGTCAGG AGTTCAAGAC CAGCCTGACC AACATGGAGA AACCCCTCT							23543
CTACTAAAAA TACAAAATTA GCTGGGCGTG GTGGCATATG CCTGTAATCC CAGCTACTCG							23603
GGAGGCTGAG GCAGGAGAAT CTTTGAACC CGGGAGGCAG AGGTTGCGAT GAGCCTAGAT							23663
CGTGCCATTG CACTCCAGCC TGGGCAACAA GAGCAAAACT CGGTCTCAAA AAAAAA							23723
AAAAAGTGA ATTAACCAAA GGCATTAGCT TAATAATTTA ATACTGTTTT TAAGTAGGGC							23783
GGGGGTGGC TGAAGAGAT CTGTGTAAT GAGGGAATCT GACATTTAAG CTTTCATCAGC							23843
ATCATAGCAA ATCTGCTTCT GGAAGGAACT CAATAAATAT TAGTTGGAGG GGGGGAGAGA							23903
GTGAGGGGTG GACTAGGACC AGTTTTAGCC CTTGTCTTTA ATCCCTTTTCT CTGCCACTAA							23963

TAAGGATCTT	AGCAGTGGTT	ATAAAAGTGG	CCTAGGTTCT	AGATAATAAG	ATACAACAGG	24023
CCAGGCACAG	TGGCTCATGC	CTATAATCCC	AGCACTTTGG	GAGGGCAAGG	CGAGTGTCTC	24083
ACTTGAGATC	AGGAGTTCAA	GACCAGCCTG	GCCAGCATGG	CGATACTCTG	TCTCTACTAA	24143
AAAAAATACA	AAAATTAGCC	AGGCATGGTG	GCATGCACCT	GTAATCCCAG	CTACTCGTGA	24203
GCCTGAGGCA	GAAGAATCGC	TTGAAACCAG	GAGGTGTAGG	CTGCAGTGAG	CTGAGATCGC	24263
ACCACTGCAC	TCCAGCCCTG	GCGACAGAAT	GAGACTTTGT	CTCAAAAAAA	GAAAAAGATA	24323
CAACAGGCTA	CCCTTATGTG	CTCACCTTTC	ACTGTTGATT	ACTAGCTATA	AAGTCCTATA	24383
AAGTTCTTTG	GTCAAGAACC	TTGACAACAC	TAAGAGGGAT	TTGCTTTGAG	AGGTTACTGT	24443
CAGAGTCTGT	TTCATATATA	TACATATACA	TGTATATATG	TATCTATATC	CAGGCTTGCG	24503
CAGGGTTCCC	TCAGACTTTC	CAGTGCACCT	GGGAGATGTT	AGGTCAATAT	CAACTTTCCC	24563
TGGATTACAG	TTCAACCCCT	TCTGATGTAA	AAAAAAGAAA	AAAAAAGAAA	GAAATCCCTT	24623
TCCCCTTGGA	GCACTCAAGT	TTCACCAGGT	GGGGCTTTCC	AAGTTGGGGG	TTCTCCAAGG	24683
TCATTGGGAT	TGCTTTCACA	TCCATTTGCT	ATGTACCTTC	CCTATGATGG	CTGGGAGTGG	24743
TCAACATCAA	AACTAGGAAA	GCTACTGCCC	AAGGATGTCC	TTACCTCTAT	TCTGAAATGT	24803
GCAATAAGTG	TGATTAAAGA	GATTGCCTGT	TCTACCTATC	CACACTCTCG	CTTTCAACTG	24863
TAACTTTCTT	TTTTTCTTTT	TTTCTTTTTT	TCTTTTTTTT	TGAAACGGAG	TCTCGCTCTG	24923
TCGCCCAGGC	TAGAGTGCAG	TGGCACGATC	TCAGCTCACT	GCAAGCTCTG	CCTCCCGGGT	24983
TCACGCCATT	CTCCTGCCTC	ACCCTCCCAA	GCAGCTGGGA	CTACAGGCGC	CTGCCACCAT	25043
GCCCAGCTAA	TTTTTTGTAT	TTTTAGTAGA	GACGGGGTTT	CACCGTGTTA	GCCAGGATGG	25103
TCTCGATCTC	CTGAACCTGT	GATCCGCCCC	CCTCAGCCTC	CCAAAGTGCT	GGGATTACAG	25163
GCGTGAGCCA	TCGCACCCGG	CTCAACTGTA	ACTTTCTATA	CTGGTTCATC	TTCCCTTGTA	25223
ATGTTACTAG	AGCTTTTGAA	GTTTTGGCTA	TGGATTATTT	CTCATTTATA	CATTAGATTT	25283
CAGATTAGTT	CCAAATTGAT	GCCCACAGCT	TAGGGTCTCT	TCCTAAATTG	TATATTGTAG	25343
ACAGCTGCAG	AAGTGGGTGC	CAATAGGGGA	ACTAGTTTAT	ACTTTCATCA	ACTTAGGACC	25403
CACACTTGTT	GATAAAGAAC	AAAGGTCAAG	AGTTATGACT	ACTGATTCCA	CAACTGATTG	25463
AGAAGTTGGA	GATAACCCCG	TGACCTCTGC	CATCCAGAGT	CTTTCAGGCA	TCTTTGAAGG	25523
ATGAAGAAAT	GCTATTTTAA	TTTTGGAGGT	TTCTCTATCA	GTGCTTAGGA	TCATGGGAAT	25583
CTGTGCTGCC	ATGAGGCCAA	AATTAAGTCC	AAAACATCTA	CTGGTTCAG	GATTAACATG	25643
GAAGAACCCT	AGGTGGTGCC	CACATGTTCT	GATCCATCCT	GCAAAATAGA	CATGCTGCAC	25703
TAACAGGAAA	AGTGCAGGCA	GCCTACCCAG	TTGGATAACC	TGCAAGATTA	TAGTTTCAAG	25763
TAATCTAACC	ATTTCTCACA	AGGCCCTATT	CTGTGACTGA	AACATACAAG	AATCTGCATT	25823
TGGCCTTCTA	AGGCAGGGCC	CAGCCAAGGA	GACCATATTC	AGGACAGAAA	TTCAAGACTA	25883
CTATGGAAC	GGAGTGCTTG	GCAGGGAAGA	CAGAGTCAAG	GACTGCCAAC	TGAGCCAATA	25943
CAGCAGGCTT	ACACAGGAAC	CCAGGGCCTA	GCCCTACAAC	AATTATTGGG	TCTATTCACT	26003
GTAAGTTTAA	ATTTTCAGGCT	CCACTGAAAG	AGTAAGCTAA	GATTCTCTGG	ACTTTCTGTC	26063
TCTCTCACAG	TTGGCTCAGA	AATGAGAACT	GGTCAGGCCA	GGCATGGTGG	CTTACACCTG	26123
GAATCCCAGC	ACTTTGGGAG	GCCGAAGTGG	GAGGGTCACT	TGAGGCCAGG	AGTTCAGGAC	26183
CAGCTTAGGC	AACAAAGTGA	GATACCCCTT	GACCCCTTCT	CTACAAAAAT	AAATTTTAAA	26243
AATTAGCCAA	ATGTGGTGGT	GTATACTTAC	AGTCCCAGCT	ACTCAGGAGG	CTGAGGCAGG	26303
GGGATTGCTT	GAGCCCAGGA	ATTCAAGGCT	GCAGTGAGCT	ATGATTTTAC	CACTGCACCT	26363
CTGGCTGGGC	AACAGAGCGA	GACCCTGTCT	CAAAGCAAAA	AGAAAAAGAA	ACTAGAACTA	26423
GCCTAAGTTT	GTGGGAGGAG	GTCATCATCG	TCTTTAGCCG	TGAATGGTTA	TTATAGAGGA	26483
CAGAAATTGA	CATTAGCCCA	AAAAGCTTGT	GGTCTTTGCT	GGAACCTCTAC	TTAATCTTGA	26543
GCAAAATGTG	ACACCACTCA	ATGGGAGAGG	AGAGAAGTAA	GCTGTTTGAT	GTATAGGGGA	26603
AAACTAGAGG	CCTGGAACTG	AATATGCATC	CCATGACAGG	GAGAATAGGA	GATTCCGAGT	26663
TAAGAAGGAG	AGGAGGTCAG	TACTGCTGTT	CAGAGATTTT	TTTTATGTAA	CTCTTGAGAA	26723
GCAAACTAC	TTTTGTTCTG	TTTGTAATA	TACTTCAAAA	CAAACTTCAT	ATATTCAAAT	26783
TGTTTCATGTC	CTGAAATAAT	TAGGTAATGT	TTTTTTCTCT	ATAG GAA ATG AAT CCT		26839
				Glu Met Asn Pro		
				85		
CCT GAT AAC ATC AAG GAT ACA AAA AGT GAC ATC ATA TTC TTT CAG AGA						26887
Pro Asp Asn Ile Lys Asp Thr Lys Ser Asp Ile Ile Phe Phe Gln Arg						
90		95		100		
AGT GTC CCA GGA CAT GAT AAT AAG ATG CAA TTT GAA TCT TCA TCA TAC						26935
Ser Val Pro Gly His Asp Asn Lys Met Gln Phe Glu Ser Ser Ser Tyr						
105		110		115		120
GAA GGA TAC TTT CTA GCT TGT GAA AAA GAG AGA GAC CTT TTT AAA CTC						26983
Glu Gly Tyr Phe Leu Ala Cys Glu Lys Glu Arg Asp Leu Phe Lys Leu						
125		130		135		
ATT TTG AAA AAA GAG GAT GAA TTG GGG GAT AGA TCT ATA ATG TTC ACT						27031
Ile Leu Lys Lys Glu Asp Glu Leu Gly Asp Arg Ser Ile Met Phe Thr						
140		145		150		
GTT CAA AAC GAA GAC T AGCTATTAAA ATTTTCATGCC GGGCGCAGTG GCTCACGCCT						27087
Val Gln Asn Glu Asp						
155						
GTAATCCCAG CCCTTTGGGA GGCTGAGGCG GGCAGATCAC CAGAGGTCAG GTGTTCAAGA						27147
CCAGCCTGAC CAACATGGTG AAACCTCATC TCTACTAAAA ATACAAAAAA TTAGCTGAGT						27207

GTAGTGACCC	ATGCCCTCAA	TCCCAGCTAC	TCAAGAGGCT	GAGGCAGGAG	AATCACTTGC	27267
ACTCCGGAGG	TGGAGGTTGT	GGTGAGCCGA	GATTGCACCA	TTGCGCTCTA	GCCTGGGCAA	27327
CAACAGCAAA	ACTCCATCTC	AAAAAATAAA	ATAAATAAAT	AAACAAATAA	AAAATTCATA	27387
ATGTGAACTG	TCTGAATTTT	TATGTTTAGA	AAGATTATGA	GATTATTAGT	CTATAATTGT	27447
AATGGTGAAA	TAAAATAAAT	ACCAGTCTTG	AAAAACATCA	TTAAGAAATG	AATGAACCTT	27507
CACAAAAGCA	AACAAACAGA	CTTTCCCTTA	TTTAAGTGAA	TAAAATAAAA	TAAAATAAAA	27567
TAATGTTTAA	AAAATTCATA	GTTTGAAAC	ATTCTACATT	GTTAATTGGC	ATATTAATTA	27627
TACTTAATAT	AATTATTTTT	AAATCTTTTG	GGTTATTAGT	CCTAATGACA	AAAGATATTG	27687
ATATTTGAAC	TTTCTAATTT	TTAAGAATAT	CGTTAAACCA	TCAATATTTT	TATAAGGAGG	27747
CCACTTCACT	TGACAAATTT	CTGAATTTCC	TCCAAAGTCA	GTATATTTTT	AAAATTCAGT	27807
TTGATCCTGA	ATCCAGCAAT	ATATAAAAGG	GATTATATAC	TCTGGCCAAC	TGACATTCAT	27867
CCTAGGAATG	CAAAGATGGT	TTAATATCCT	AAAATCAATT	AACATAACAT	ACTATATTAA	27927
TAAAGTATCA	AAACAGTATT	CTCATCTTTT	TTTCTTTTTT	CACAATTCCT	TGGTTACACT	27987
ATCATCTCAA	TAGATGCAGA	AAAAGCATTT	GACAAAATCC	AATTCATAAT	AAAAATTCTC	28047
AACTTGAAA	GAGAACATCA	TAAAGGCATC	TATGAAAAAC	CTACAGCTAA	TATCATACTT	28107
AACGATGAAA	AACCTGAATTA	TTTTACCTTA	AGATCAAGAA	TAATGCAAGC	ATGTCAGCTC	28167
TTGCAACTTC	TATTCAACAT	TGTACTGGAG	GTTCTAGCCA	GAGCAACCAT	ACAATAAATA	28227
AAAATAAAAG	GCACCCAGAT	TAGAAAGGAA	GTCTTTATTT	GCAGACAACA	TGGTTCTTTA	28287
TGCAGAAAAC	CGTCAGGAAT	ACACACACAT	GTTAGAACTA	ATAAGTTCAG	CAAGGTTGCA	28347
GGTTGCAATA	TCAATATGCA	AAAATACATT	GAAGGCTGGG	CTCAGTGGAG	ATGGCATGTA	28407
CCTTTTCGTCC	CAGCTACTTG	GGAGGCTGAG	GTAGGAGGAT	CACTTGAGGT	GAGGAGTTTG	28467
AGGCTATAGT	GCAATGTGAT	CTTGCCTGTG	AATAGCCACT	GCACTCGAGC	CTAGGCAACA	28527
AAGTGAGACC	CCGTCTCCAA	AAAAAAAAT	GGTATATTGG	TATTTCTGTA	TATGAACAAT	28587
GAATGATCTG	AAAACAAGAA	AATTCCATTC	ACGATGGTAT	TAAAAAATA	AAATACAAAT	28647
AAATTTAGCA	AAATAATTAT	AAAACCTGTA	CATCGAAAAT	TTCAAAGCAC	TCTGAGGGAA	28707
ATTAAAGATG	ATCTAAATAA	TTGGAGAGAC	ACTCTATGAT	CACTGATTGG	AAAATTCATT	28767
CAATATTGTT	AAGATAACAA	TTGTCCCCAA	ATTGATGCAT	GCATTCAATT	TAGTCTTCAT	28827
CAAAATTCCA	GCAGGGTTTT	TGCAGAAATT	GACAAGCTGT	ACCCAAAATG	TATATGAAAA	28887
TGAAAAGACC	CAGAAGAGCA	AATAATTTTT	TAAAAACAAA	GTTGGAAAAC	TTTTACTTCC	28947
TAATTTTAAA	ACTTACTATA	AACCTAAAGT	TATCAAGACC	ATTTAGT		28994

(15) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 10 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(v) FRAGMENT TYPE: N-terminal fragment

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

Tyr Phe Gly Lys Leu Glu Ser Lys Leu Ser
1 5 10

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CCATCCTAAT ACGACTCACT ATAGGGC

27

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 28 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CTATAGGGCA CGCGTGGT

18

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TTCCTCTTCC CGAAGCTGTG TAGACTGC

28

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GTAAGTTTTC ACCTTCCAAC TGTAGAGTCC

30

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGGATCAAGT CGTGATCAGA AGCAGCACAC

30

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CCTGGCTGCC AACTCTGGCT GCTAAAGCGG

30

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTATTGTCAA TAAATTTTCAT TGCCACAAAG TTG

33

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AAGATGGCTG CTGAACCACT AGAAGACAAT TGC

33

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TCCTTGGTCA ATGAAGAGAA CTTGGTC

27

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC

33

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

GCCAGCCTAG AGGTATGGCT GTAACATCT C

31

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

GGCATGAAAT TTTAATAGCT AGTCTTCGTT TTG

33

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

GTGACATCAT ATTCTTTCAG AGAAGTGTCC

30

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAATTTGAA TCTTCATCAT ACGAAGGATA C

31

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 31 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TCCGAAGCTT AAGATGGCTG CTGAACCACT A

31

(2) INFORMATION FOR SEQ ID NO:33:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 32 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GGAAATAATT TTGTTCTCAC AGGAGAGAGT TG

32

(2) INFORMATION FOR SEQ ID NO:34:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 37 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

ATGTAGCGGC CGCGGCATGA AATTTTAATA GCTAGTC

37

(2) INFORMATION FOR SEQ ID NO:35:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

CCTGGAATCA GATTACTTTG GCAAGCTTGA ATC

33